OCTD

Oxford Centre for Technology and Development

Using Data Analysis and Artificial Intelligence/Machine Learning Techniques for Humanitarian Cash Transfer and Digital Identification

Report

07.05.2024

SECTION 1: HUMANITARIAN CASH TRANSFER AND OVERCOMING DATA SCARCITY

1.1. Market Assessment in Emergency Contexts

In disaster and war struct contexts, it is both vital and difficult to assess market dynamics and value to determine the most effective ways to allocate cash transfers and measure their reach as well as effectiveness. Different qualitative and quantitative market analysis tools have been established to guide researchers and decision-makers. EMMA, MIFIRA, and RAM are exemplary tools.

EMMA stands for Emergency Market Mapping and Analysis. EMMA framework is used in sudden-onset emergencies to identify market systems. It provides researchers a tentative ten-step framework involving population and need targeting, selecting critical market system(s), household profiling, fieldwork preparation, fieldwork conduct, mapping, identifying shortfalls & access constraints, identifying market capacity & performance, defining response strategy, and communicating results (Albu, 2010, p.14). According to the World Food Program analysis (Kamara, 2013, p.5) the EMMA framework is flexible and comprehensive, yet heavily relies on qualitative information gathered via extensive fieldwork.

MIFIRA stands for Market Information and Food Insecurity Response Analysis. MIFIRA framework provides sets of questions in the form of a decision tree to guide researchers to appropriate tools and methods to assess the circumstances of food markets and behaviours of the key stakeholders in these markets (Barrett et al., 2009, p.156). According to the WFP analysis (Kamara, 2013, p.5), the MIFIRA framework relies on predetermined sets of questions and quality secondary data which cannot be adapted to rapid-onset emergency contexts.

RAM stands for Rapid Assessment for Markets. RAM framework is used for identifying key commodity markets in the immediate aftermath of disasters (ICRC, 2014, p.5). RAM is composed of five steps involving defining the characteristics of the affected area before disaster and impact(s) of disaster, collecting data on key informants and traders via interviews and focus groups, utilising a decision tree to determine the capacity of marketplaces to respond demand shifts, reporting, and developing tools to monitor market changes (ICRC, 2014, p.6).

The main advantages of these tools are comprehensiveness and adaptability in different contexts to develop novel assessment strategies considering the divergent characteristics of emergency, governance, and stakeholder networks in structuring and adapting markets before, during, and after disaster events. Considering the definition of market is a difficult task in itself, it is also important to differentiate and account for formal and informal markets before and after disasters. In contexts where formal market relations are inadequate to capture the consumption and trade relations, the adaptability of these frameworks should be utilised and novel digital tools should account for socioeconomic-demographic-cultural determinants affecting the consumption habits, ways of "doing business", and community networks.

To assess the state of local markets, price discussions, market trends, geographic distribution of markets/shops, and market access can be conceptualised and operationalised.

Within the literature, Cavallo and colleagues measure "aggregate behaviour of prices and product availability" (Cavallo et al., 2014, p.22) in the aftermath of the 2010 earthquake in Chile and the 2011 earthquake in Japan. They create daily price indices for different goods to quantify "the degree of supply disruption with an index of product availability, which tracks the number of goods (supermarket goods such as food, beverages, etc.) that are available for purchase over time" (Cavallo et al., 2014, p.2). Price and category indexes at each goods category are calculated using the Billion Prices Project (BPP) dataset established by MIT. Moreover, they estimated the "hazard rate" (Cavallo et al., 2014, p.18) of price changes to analyse whether "there is evidence that the hazard rate of stockout decreases with time after the disaster, while the hazard rate of price changes increases" (Cavallo et al., 2014, p.18). Their results suggest that "immediate effects of the natural disaster were mostly in product availability rather than prices" (Cavallo et al., 2014, p.18).

When it comes to analysing market trends, Henein uses Generalised Autoregressive Conditional Heteroskedastic (GARCH) modelling and ordinary least square regression (OLS) to understand "the factors behind fluctuations in returns and volatility of specific agricultural commodities" (Henein, 2023, p. 15) in 46 countries. In this context, the GARCH model is used to quantify "conditional variance" or volatility of "each commodity returns" by taking their past variances into account (Henein, 2023, p. 19). He utilises "event study procedure" (Henein, 2023, p.15) to identify the events of interest. To run these models, Henein uses Bloomberg Agriculture Subindex Total (which tracks the performance of various agriculture commodities) and FAOSTAT databases. According to the results, certain factors such as "substitutability, storability, and geographical concentration of production" (Henein, 2023, p. 33) affect the reactions and consequently prices and stocks differently. Moreover, the results suggest that short-term natural disasters do not affect commodity prices in line with the conventional economic logic. Rather, markets are attentive to the long-term impacts of disasters which may be determined by the amount of time passed to process and diffuse disaster related information (Henein, 2023, p. 33). In a longer timeframe, the prices are suggested to "revert to their mean" (Henein, 2023, p. 33). However, there may be errors regarding causality and elevated noise levels (Henein, 2023, p. 34).

Cavallo's and Henein's studies provide important tools and datasets to observe prince changes comparatively. An important downside of these studies, however, is that the datasets' focus is quite narrow.

The geographic distribution of markets/shops and their accessibility may be important indicators to assess price and stock changes. Esmalian and colleagues examine differing access to grocery stores for socially vulnerable populations during different phases of the disaster in Harris County, Texas using population facility networks (Esmelian, et al., 2022, p.1). To understand divergences in access, examine the dimensions of access alongside variacions and influencing factors . To operationalize this process they established the population-facility network models of the study area on which nodes represent traffic analysis zone (TAZ), facility nodes represent individual POIs, and the links represent trips (Esmelian, et al., 2022, p.4). Thus, three network models were created. One with unweighted links, with weighted links representing trip duration, with weighted trip distance links (Esmelian, et al., 2022, p.4). Mobility data from StreetLight, sociodemographic data from American Community Survey of the U.S. Census Bureau, and Food Access Research Atlas were utilised. Anonymized data from cell phones and GPS devices were used to create travel metrics, such as duration and distance (Esmelian, et al., 2022, p.3). Location of the shops were gathered from SafeGraph. Time-series clustering algorithms on the access indicators were used to identify the spatial and temporal variations in access (Esmelian, et al., 2022, p.6). To handle missing data in time series, the Kalman imputation method was utilised (Esmelian, et al., 2022, p.5). According to the results, the importance of identifying access to shops rely on three fundamental dimensions that are "redundancy, proximity, and rapidity" (Esmelian, et al., 2022, p.12). The conventional measures such as location in a food desert and the number of available stores in an area, is not sufficient for understanding access in emergency contexts (Esmelian, et al., 2022, p.12).

The analysis of market activities should also incorporate how households interact with different local market mechanisms. A study by Huddleston and Wood reveal what kinds of market mechanisms emerge and how households derive their sustenance in conflict ridden contexts in a prolonged time frame such as Yemen. They argue that functional and illegal markets are not synonymous. Functional markets emerge where prolonged uncertainty weakens regulated market relations and central authority, replacing with "new local elite networks", commodities, and trade rules (Huddleston & Wood, 2021, p. 208). To understand these replacements they conduct surveys, focus groups, and key informant interviews.

When it comes to the informal markets, the World Bank created The Informal Sector Enterprise Surveys to provide a snapshot of the state of the informal sector in countries from East Asia & Pacific, Latin America & Caribbean, Middle East & North Africa, South Asia, and Sub-Saharan Africa. In addition, UNDP provides Informal Economy Data Explorer.

In emergency contexts, it is important to understand human mobility patterns to observe shifts in demand and prices alongside the distribution of population to determine the target and size of aid. Cumbane proposes various data sources such as LIDAR and anonymized mobile phone CDRs to "estimate spatial distribution of displaced population" (Cumbane, 2022, p.44). The estimation method involves "Voronoi tessellation of the study area, estimation of mobile phone users home location (at cell-tower level) before and after disaster, neighbourhood home location assignment before and after disaster, estimation of displaced mobile phone users, Scaling up the displaced mobile phone users to actual population flow, and validation" (Cumbane, 2022, p.23). The results indicate that, If provided in a real-time way, CDR can be used to derive the "near real-time displacement matrix" (Cumbane, 2022, p.41).

Targeting populations with greatest need is one of the most challenging dimensions of humanitarian aid. Aiken and colleagues first estimated relative wealth by applying ML to satellite imagery. After, average daily consumption of each mobile subscriber was estimated by ML via mobile phone metadata. Then, survey data regarding consumption were matched with the metadata regarding each subscriber's phone-use history. "This sample was used to train supervised machine-learning algorithms that predict wealth and consumption from phone use" (Aiken et al., 2022, p.866).

Assessing whether aid has reached the target areas especially on the community level remains somewhat hidden within the literature. Jung estimates "aid presence and density, given regional poverty" in Myanmar by linking aid location with nightlight luminosity and wealth data (Jung, 2023, p.1). He uses aid location data from web portal interfaces and administrative documents, wealth and population data from the nationally representative survey to assess long-term well-being, satellite imagery of night light to assess economic activity; conflict event and vulnerability data to assess fragility (Jung, 2023, p.5). He analyses the association between wealth and aid location through three perspectives: "presence, density, and project" (Jung, 2023, p.6). Presence analysis the presence of aid in a particular region using logistic regression with "the odds of CCD projects being present in a township, given the level of needs in that township" (Jung, 2023, p.6). Density analysis involves measuring density of projects via "least squares regression of project occurrence rate per unit area on the wealth and development indicators" (Jung, 2023, p.6). Finally, project analysis involves measuring the difference between two aid models (Jung, 2023, p.6). According to the results, "with each increment of a vulnerable population rise, the odds of aid presence in that community declines which means that nightlight luminosity is a consistent measure showing that aid tends to flow to brighter areas" (Jung, 2023, p.8).

Community sentiment can be an important tool to help targeting. A research by Ragini and colleagues developed a method to categorise disaster related Tweets via Support Vector Machine (SVM). The major impact of this model is the ability to segregate texts in real-time and geolocate them (Ragini et al., 2018).

Up until now, types of humanitarian aid were not discussed. According to a review conducted by Idris (2016), the type of aid is the "key determinant of multiplier effects" (Idris, 2016, p.2). Moreover, he states that cash transfers can have significant advantages over in-kind assistance because cash aid can allow beneficiaries to be more flexible in spending which would "distribute growth effects more widely across sectors" (Idris, 2016, p.3). However, he emphasises that the effectiveness of cash transfers require "functioning markets" (Idris, 2016, p.3) to meet demand.

Another review evaluates the effectiveness of cash transfers by analysing multipliers mainly by Social Accounting Matrix (SAM), Computable General Equilibrium (CGE), Local Economy-Wide Impact Evaluation (LEWIE), and econometric techniques. According to Gassmann and colleagues' results, "cash transfers can influence growth through macro-level, meso-level, and micro-level effects through consumption, human capital investments, productive investments, shock resilience and labour supply" (Gassmann et al., 2023, p.27).

The digital mediums of cash transfer, tools to target, distribute, and evaluate them necessitate solid data protocols and privacy frameworks. The United Nations ecosystem provides general frameworks for governing, managing, and analysing data. For example, there are UNDRR' Data Strategy and Roadmap 2023-2027, UNDP's Digital Strategy 2022-2025, and Data Strategy of the Secretary-General 2020-2022.

There are also various data responsibility guidelines published by different institutions. For example, there are International Committee of the Red Cross's (ICRC) Handbook on Data Protection in Humanitarian Action (2020), CALP Network's Data Responsibility Toolkit: A Guide for CVA Practitioners (2021), and Inter Agency Standing Committee's Operational Guidance on Data Responsibility in Humanitarian Action (2023). The CALP Network and the Centre for Humanitarian Data also provide additional guidelines. Particularly, the CALP Network involves guidelines specifically for cash transfers alongside toolkits to design effective cash transfer programs.

An important issue in digital technologies in humanitarian aid is to involve stakeholders in designing methods, tools, and algorithms. "Society in-the-loop (SITL)" (Rahwan, 2018, p.2) algorithmic design which involves embedding societal considerations and feedback to ensure transparency and algorithmic accountability as well as regulation.

1.1.1. Suggestions for the WFP

During humanitarian emergencies, where the situation is complex due to ongoing conflict, restricted access, and economic blockades, leveraging advanced technologies and innovative methodologies needs to be tailored to address these unique challenges.

Below is more specific suggestions for the WFP:

a) They can develop microsimulation models that use household survey data to predict how changes in economic conditions, due to conflict or blockade, affect household income and needs. These models can simulate the impact of cash transfers on household welfare, allowing for adjustments in transfer amounts based on predicted needs.

b) They can use remote sensing technology to assess agricultural productivity and damage in the affected location. Satellite imagery can provide data on crop health, water stress, and land use changes. Machine learning algorithms can analyze this data to estimate food availability and prices, informing cash transfer amounts needed to ensure food security.

c) They can perform network analysis on supply chain data to identify bottlenecks and vulnerabilities in the market system. Understanding how goods move into and within the region, and how conflicts or blockades affect these movements, can help predict market shortages and price spikes, guiding the cash transfer strategy.

d) They can implement digital payment platforms for cash transfers that can operate with limited internet connectivity. Considering the frequent power outages and internet disruptions in crises-affected regions, digital wallets that sync transactions offline and update once connectivity is restored can ensure uninterrupted cash assistance.

e) They may also analyze unstructured data from local news outlets, social media, and community forums using NLP techniques to monitor the situation in real-time. This can provide early warnings of escalating conflict, market disruptions, or emerging needs, allowing for swift adjustments to cash transfer programs.

f) They can deploy AI-enabled chatbots on popular messaging platforms to collect feedback from beneficiaries about the effectiveness of cash transfers, market conditions, and barriers to accessing aid. Chatbots can provide a scalable way to engage with the community, gather insights, and adjust programs based on beneficiary feedback.

g) They can use GIS and geospatial analysis to identify safe and accessible cash distribution points. In conflict zones, ensuring the safety of beneficiaries and aid workers is paramount. Mapping conflict intensity, movement restrictions, and beneficiary locations can inform the placement of distribution points.

h) They can apply machine learning models to historical conflict data and real-time signals (e.g., social media activity, news reports) to predict changes in conflict intensity. These predictions can help in planning cash transfers around periods of lower conflict intensity, reducing risks for beneficiaries and aid workers.

1.2. METHODOLOGICAL SECTION

A number of methods exist for handling less-than-ideal datasets, including data scarcity. These methods come from statistics, economics, marketing, machine learning, biostatistics, and elsewhere. These methods can be broadly categorized into three groups depending on their core strategy or approach. The following lists the most important methods, described in the context of cash transfers:

1. Make more efficient use of existing data
    - i. Missing data interpolation
    - ii. Pooling approaches (e.g., hierarchical models)
    - iii. Regularization and Informative Priors
    - iv. Inertia
2. Augment existing data with new sources
    - i. Data Fusion: small-scale primary data, aggregate moments
    - ii. Surrogate models

iii. Online trace data

iv. Qualitative Data Enrichment

3. Reduce negative consequences associated with scarce data

i. Uncertainty quantification

The remainder of the chapter contains concrete recommendations for applying these methods in a data-scarce context. For the purposes of this chapter, "data scarcity" could have multiple interpretations: missing variables, missing observations, infrequent data collection, and highly aggregated data. It will be important to clearly link methods to the types of data scarcity problems they solve or mitigate.

Before moving into the details, it will be useful to have fixed notation to explain the methods more clearly. We will often refer to this equation

$$y_{bt} = h(\alpha_{bt}, y_{bt-m}, AMT_{BT}, X_{bt}, Z_{bt}, \epsilon_{bt})$$

The variable y_{bt} is an outcome of interest, such as consumption expenditure. The variable α_{bt} represents unit indicators, which could be at various levels of aggregation (e.g., individuals, communities, villages, etc.). Seasonality and other time trends could be captured by dependence of $h(\cdot)$ on y_{bt-m}The variable AMT_{BT} is the vector of transfer amounts. Note that the outcome for y_{bt} could depend on transfer amounts given to other people or to the same household in previous periods (see Egger et al. 2022). The matrix X_{bt} contains variables of interest subject to data scarcity, in particular prices, and the matrix Z_{bt} contains control variables. Finally ϵ_{bt} contains unobservables. Commonly, $h(\cdot)$ has a linear form:

$$y_{bt} = \alpha_{bt} + g(y_{bt-m}, \gamma, t) + f(AMT_{BT}, \Delta) + X_{bt}\beta_b Z_{bt}\delta_b + \epsilon_{bt}$$

Where $g(\cdot)$ captures autocorrelation or other time-dependence (possibly through an ARIMA structure), possibly involving parameters γ and $f(\cdot)$ captures spatial and temporal dependence in cash transfers, possibly with parameters Δ. A key contextual question is whether the transfer amounts should be jointly or individually optimised. In what follows we use the simpler paradigm of individual optimisation, and write AMT_{bt} instead of $f\left(AMT_{BT}, \Delta\right)$. However, in most cases the extension to joint optimisation would be straightforward. To simplify notation we will also group the first two terms together and suppress the dependence on arguments:

$$\alpha_{bt} + g\left(y_{bt-m}, \gamma, t\right) \equiv A$$

With these specifications we have

$$y_{bt} = A + AMT_{bt}\Delta_{bt} + X_{bt}\beta_b Z_{bt}\delta_b + \epsilon_{bt}$$

Extending from the linear specification to more general non-linear specifications for $h(\cdot)$ may not always be straightforward, but the general principles will usually apply.

More Efficient Use of Existing Data

Commonly, data are not used to their full potential. This section collects a set of tools for using data more efficiently. These tools are related to and inspired by Bayesian statistics, but in many cases can be used with traditional workflows and tools (e.g., frequentist statistics or ML tools.)

In recent years the tools associated with Big Data have grown and spread, and in some ways the tools of “small data” have been neglected by practitioners. A key pillar in tools for “small data” is Bayesian statistics. Bayesian statistics begin with a paradigm of limited information. Bayesian approaches explicitly incorporate prior knowledge, uncertainty, and the risks of making certain decisions (through Statistical Decision Theory). There are, however, some drawbacks associated with Bayesian methods: they are slow to compute, they don’t come with statistical guarantees for large data (i.e., Bayesian posterior distributions may not concentrate around the “true” parameter even asymptotically), and they are not well known and so decision makers struggle to interpret them. With those pros and cons in mind, we set out here to draw inspiration from Bayesian methods while maintaining as much flexibility as possible in implementation. In some cases there is no escaping Bayesian implementations, but new software libraries have made these easily accessible and familiar to a large group of data scientists and statisticians. On the other hand in some cases we can largely get the benefit of Bayesian methods but in a simpler or quicker way.

Missing Data Interpolation

Bayesian inference handles missing data as if it were any other unknown parameter. Combined with the fact that Bayesian methods yield an entire distribution as an estimate, this is ideal for interpolation. The reason this latter feature is helpful is because uncertainty in the interpolation will pass through into the final estimates. Other approaches, where predictions for the missing values are plugged in as if they were data, do not capture any of the uncertainty in interpolation. (Andridge and Little 2010, Kamakura and Wedel 1997; Gilula et al. 2006; Quan and Xie 2014).

Interpolation could be useful in a number of ways. Here we outline in detail two separate approaches in the context of cash transfers, which should provide enough of an example to adapt further.

1. Spatial: Interpolating y_{bt}, X_{bt}, or Z_{bt}across regions to use data from one place for a different place.
2. Temporal (Time Series): Interpolating y_{bt}, X_{bt}, or Z_{bt} across time periods, which may allow for more frequent estimates.

There are many methods that could in principle be used for interpolation, even within Bayesian approaches. Broadening the scope beyond Bayesian approaches, there are many more.

Spatial Interpolation Options

For interpolating spatially (option 1 above), a promising approach may be to adapt the Synthetic Control method from economics (Abadie, Gardeazabal 2003; Abadie et al. 2010) to the task of interpolation. In fact, Synthetic Control as a method is essentially an approach to extrapolation or interpolation (the exact term depends on context). This can be done using a Bayesian approach but is most commonly done using tools from the frequentist tradition (non-Bayesian).

A practical application of **Synthetic Control** involving data scarcity in cash transfer settings would look like the following:

1. Choose a focal area with data sparsity
2. Find nearby areas or areas with similar characteristics to serve as a "donor pool" for data.
3. Build a synthetic unit from the donor pool using a preferred model. The model is typically chosen mostly to maximize predictive accuracy in a validation sample (see Doudchenko and Imbens 2016). Since these validation samples are chosen using a cutoff in time, this method is directly applicable to temporal interpolation (see below.) Alternatively, a matrix completion method can be used which handles missing data naturally (Athey et al. 2020). Bayesian Synthetic Control is another possibility (Kim et al. 2020).
4. Use the chosen approach from the previous step to impute missing data from the focal area. The key difference for this application is that there may be no treatment effect to estimate. Instead, the focal unit might be under the same conditions as all units in the donor pool.
5. Use the interpolated data as if it were normal data, either with imputed data plugged in without accounting for uncertainty, or run as jointly with the above methods where applicable. The matrix completion method is a joint analysis of both imputation and the final model, as is the Bayesian Synthetic Control. Again, the difference between Synthetic Control and Synthetic Interpolation is that the observed outcome is omitted in the proposed Synthetic Interpolation (meaning, very simply, the final step is not necessary).

See Abadie (2021) for more practical details on Synthetic Control. To implement the Synthetic Interpolation, it may be possible to adapt Synthetic Control programs by simply entering zero for each "observed" value. Then the treatment effects would be the negative of the interpolated values. This may make this method very easy to implement.

A second option for interpolating across regions is **Spatial Interpolation,** specifically **kriging**, from the field of spatial statistics. The structure is very similar to the Synthetic Interpolation described above. The basic idea of kriging is to use a set of sampled points across a space to estimate the value of a nearby unknown point. The process as applied to cash transfers would be as follows:

1. Choose a focal area with data sparsity.
2. Find areas with observed values of the variable to be interpolated to serve as a "control points" or "sample points" in the terminology of spatial interpolation.
3. Build an interpolation model over the control points using a preferred spatial model. The most widely used model is the Gaussian Process without covariates (Gelfand and Schliep 2016), or Bayesian regression-based kriging using integrated nested Laplace approximations (INLA) with covariates (Wang and Furrer 2019). If covariates are used to aid interpolation (say with y_{bt}, they should be distinct from covariates used in the outcome model (i.e., they should not be part of X_{bt} or Z_{bt}).
4. Use the spatial model to impute values in the focal area.
5. Use the data as normal, either with imputed data plugged in without accounting for uncertainty, or run jointly with the above methods to account for uncertainty.

See Wang and Furrer (2019) for more practical details on estimation when using covariates, or Watson (2016) for a practitioners guide without covariates. Both use Stan, a probabilistic programming language, to specify the Bayesian models and draw from the posterior distribution.

Temporal (Time Series) Interpolation Options

Interpolation over time could in principle be used to estimate data values before they are fully collected, or between data collection periods. In essence this may allow a higher degree of frequency of use from the same data.

The techniques for temporal interpolation are very similar to those for spatial interpolation, and would require minimal adaptation. In fact, the Synthetic Interpolation method discussed above is directly applicable, since it already relies on a temporal projection to impute missing data. On the other hand, spatial interpolation methods are often cross-sectional. However, there are important spatio-temporal variants (van Zoest et al. 2020), which include a time trend component. Regression-kriging could be applied by simply adding spatial and temporal covariates. For spatial forecasting outside of kriging, ARIMA has been used in epidemiology to forecast dengue (Thiruchelvam et al., 2021), and may apply in the cash transfer context to impute missing data in y_{bt} (or other target variables).

Pooling Approaches (Hierarchical Models)

When panel data are unbalanced (some units have more observations than others), it is common to use hierarchical models which have a major advantage: they pool information across units. In the above section, the reliance on data-rich units to interpolate data to the data-scarce unit is done as a means of filling in missing data. In this section, we take the data as they are and use hierarchical models to allow information from data-rich units to inform parameter estimates in data-scarce units.

A model with high potential for application to data scarcity is the **Gaussian Process Dynamic Heterogeneity (GPDH)** model of Dew et al. (2020). This model is a hierarchical Gaussian Process, where parameters of a given unit vary over time (β_{bt}, note the t index). In the cash transfer context, the individual unit is a market, or location, or even a product within a basket of goods. An outline of the application of the GPDH is as follows:

1. Assemble a panel dataset of all covariates (X_{bt}, Z_{bt}) and outcomes (y_{bt}) .
2. The lower level model relates outcomes (y_{bt}) to covariates. The model must be formulated as a Bayesian regression.
3. The upper level model relates the parameters (Δ_{bt}, β_{bt}) across different units to one another and across time.
4. If the panel is balanced, the GPDH as specified in 2 and 3 applies directly.
5. If the panel is unbalanced, the missing values for each panel would be treated as unknowns to be estimated. This could be accommodated with some customization of the GPDH program (which is written in Stan). The customization would depend on the exact dataset and context.

While pooling is commonly used in statistical modeling, it may be underexploited in the sense that typical datasets limit the number of units included in the sample. The GPDH in particular could use information over long periods of time within and across many units to produce better and more sensible MMM parameter estimates.

Regularization and Informative Priors

In machine learning, regularization is a technique used to avoid overfitting. Overfitting occurs when a model is too closely fit to the training data, and does not generalize well to new data. Regularization is a technique that can be used to avoid overfitting by adding a penalty to the error function. The penalty encourages the model to find a simpler solution, which is less likely to overfit the data. Bayesian priors can be interpreted as a type of regularization because they "exploit a sensitivity–stability trade-off: they stabilize estimates and predictions by making fitted models less sensitive to certain details of the data." (Gelman and Shalizi 2013). Some priors influence the estimates more than others.

Regularization for Cash Transfers

Regularization could help in data scarce contexts by reducing what is required of the data. For instance, a LASSO penalty parameter could be used to remove terms from the outcome equation. With fewer terms, the data points could estimate parameters with more efficiency. Very often this increased efficiency comes with minimal reduction in predictive performance, and in fact predictive performance often increases. A linear model with a regularization penalty would have the same form as above:

$$y_{bt} = A + AMT_{bt}\Delta_{bt} + X_{bt}\beta_b + Z_{bt}\delta_b + \epsilon_{bt}$$

However, to estimate the model, Penalized Maximum Likelihood Estimation could be used:

$$\left(\hat{\Delta}_b, \hat{\beta}_b, \hat{\delta}_b\right) = argmax_{\Delta_b, \beta_b, \delta_b} l\left(\Delta_b, \beta_b, \delta_b \,|\, AMT_{bt}, X_{bt}, Z_{bt}\right) - \lambda \left|\left|\left(\Delta_b, \beta_b, \delta_b\right)\right|\right|_1$$

The $l\left(\Delta_b, \beta_b, \delta_b \,|\, AMT_{bt}, X_{bt}, Z_{bt}\right)$ is the likelihood function implied by the model. This will depend on assumptions about ϵ_{bt} and the functional form used.

The tuning parameter λ is most often set using cross-validation, and it would make sense to use out-of-sample prediction performance of sales in data scarce regions as the performance criterion during cross-validation.

Regularized regressions are not new, but they are worth highlighting here because they may be particularly well suited to data scarce situations. Estimates from scarce data tend to be noisier, and regularization is in essence a noise-mitigation technique.

Informative Priors

Priors afford Bayesian models more stability and in some cases can increase accuracy and generalization if the prior reduces the sensitivity of estimates to noise in the data. The distinction between "informative" and "non-informative" priors is tricky and controversial, but in general informative priors are priors that are based on information that is known about the distribution of the data. This information can be derived from past data, expert knowledge, experiments, or other sources.

Choosing a sensible informative prior can be quite challenging. However, academic literature in economics has produced a number of candidates for prior distributions that can be used directly for cash transfer models. For example Egger et al. (2022) conduct a large scale cash transfer program and provide estimates of both micro-level outcomes as well as equilibrium effects. The outcomes include consumption expenditure increases, welfare estimates, and economic multipliers. Great effort has been made to obtain experimental evidence. Similar experimental evidence could be used to put informative priors on solid scientific footing, and in fact would probably improve estimates across all contexts (with or without data scarcity).

When past research is limited for a given variable or data context, expert judgement can be used to create priors. A number of methods exist for prior elicitation. The most natural form of elicitation is often to ask for a series of quantiles corresponding to a probability distribution. These quantiles can then be used to define a density or mass function from a parametric family, or even a quantile parameterized distribution (Keelin and Powley 2011, see also Keelin 2016 and Keelin and Howard 2021). Enumerators from country offices are likely to have more knowledge than can be easily quantified, and building that knowledge into informative priors may improve accuracy substantially.

Non-expert judgement is also feasible to incorporate into priors, and Large Language Models make this process potentially very scalable. The design in Chopra and Haaland (2023) could easily accommodate the collection of relevant information for market analysis. Ashwin et al. (2023) also describe a method for using custom AI models to *analyse* interview transcripts, while noting that using LLMs has high potential to introduce bias. The bias in Ashwin et al. appears to apply mostly to high-level, abstract concepts, which may mean that LLMs *can* be used effectively for more objective data, such as price information.

Informative priors have a major advantage in that they are simple to implement. Bayesian models already require priors, so making those priors informative involves very little modification to the estimation procedure. If the use-case is non-Bayesian, then empirical Bayes methods or penalized MLE could be used. As an example, suppose the informative prior distribution on a given coefficient (β_{bt}) is Normal with mean μ_{bt} and standard deviation σ_{bt}, and the MLE estimate is $\hat{\beta}_{bt}$ with standard error s_{bt}. Then one common form for pulling estimates closer to the prior is applying "shrinkage" after MLE estimates are obtained is

$$\beta^*_{bt} = \left(\frac{\sigma_{bt}}{\sigma_{bt} + s_{bt}}\right)\hat{\beta} + \left(\frac{s_{bt}}{\sigma_{bt} + s_{bt}}\right)\mu_{bt}$$

This can also be applied to non-Bayesian forms of hierarchical modelling, where the prior is replaced by the distribution of estimated β_{dt} for $d \neq b$, and the analogous σ_{dt}.

There is, of course, risk in using informative priors, because the priors might be worse estimates for a particular area. In machine learning terms there is a risk of "underfitting" the data scarce region. The best use cases for using informative priors are: when estimates are extremely noisy and/or when there is a high degree of confidence in the prior, such as with the experimentally determined priors mentioned above or in a setting where other units are known to be very similar to the data scarce unit (or when the risk of being wrong about that assumption is acceptable).

Inertia

When parameter estimates are the goal of an analysis project, an important question is how "slowly" the estimates (β_b) change over time. Depending on the answer to this question, as well as the intended use of the model outputs, there may be a reduced (or increased) need for frequent model estimation rounds. For example, under the extreme assumption that one parameter (say, β^1_{bt}) never changes, just a single estimation would be sufficient to identify that parameter forever. On the one hand, this assumption is unrealistic, but on the other hand, it's plausible that β^1_{bt} changes slowly enough that an estimate from a previous year could be used with little negative impact. An advantage of Bayesian methods is that they are founded upon a mechanism for updating given new information. That would, in principle, allow for a model to be updated as often as new data becomes available for *any* variable in the model.

In practice, applying Bayesian updates to the model may not always be straightforward. For Bayesian regression with conjugate Normal priors, there are closed forms for the updates (see Rossi et al. 2005). Alternatively, the model could be re-run, with all data from all previous MMM runs included. Sequential Monte Carlo (SMC) could also be used. **SMC** is unfortunately not implemented in Stan, however, it is implemented in Turing.jl, a probabilistic programming library in Julia. Finally, an approximation of the posterior from the Exponential Family of distributions could be used as the new prior. The final solution, which places ease of implementation above theoretical justification would proceed as follows:

1. Compute the Empirical Cumulative Distribution Function (ECDF) from the MCMC draws of the previous model. Then, for a list of Exponential Family distributions (Normal, Exponential, Gamma, Beta, among others), do the following:
    a. Use Maximum Likelihood Estimation to find parameters of the distribution under consideration.
    b. Simulate data from the distribution with the parameters chosen by MLE, get the ECDF and save the Kolmogorov-Smirnov distance between the Exponential Family distribution and the ECDF of the past MCMC draws.

2. The best fitting model will have the lowest Kolmogorov-Smirnov distance. An alternative is the Kullback-Leibler divergence, which would require estimating a continuous density for the previous MCMC draws.

As can be seen, there are a wide variety of approaches for using previous Bayesian posteriors for new models. The best approach in principle is SMC, but implementation may be somewhat challenging and may require customization.

An interesting alternative to the above approaches, which in theory is both easy to implement and extremely flexible, is the use of a Metalog distribution (Keelin 2016, Keelin and Howard 2021), which is a quantile-parameterized distribution (Keelin and Powley 2011). The quantiles of posterior draws from previous analysis could be used to define new Inverse CDFs (ICDFs). Stan can accommodate priors defined in terms of ICDFs (see https://github.com/bgoodri/StanCon2020).

Just as with informative priors, inertia can be accommodated inside an empirical Bayes framework, as long as there is an accompanying estimate of the variation around parameter estimates (such as a standard error or posterior variance).

Validation

Fundamentally, since cash transfers are interventions meant to improve specific outcomes, they should be analysed with causal models, or in more practical terms, it should be the case that if cash transfers were increased in line with model guidance, the predicted effects would be realized (all else held equal). The important point is that if the relationship between cash transfers and outcomes has confounds, models that fail to capture those confounds will be unreliable decision aids.

Holistic model validation is not particularly straightforward with a large number of variables and parameters. It is more feasible to validate components of high importance using a field experiment. If holistic validation is desirable, one possibility is to compare two decision policies arising different models estimated on the same data in terms of the recommendations they provide. One model "arm A" would produce recommendations for a set of units, while a second model "arm B", would produce recommendations for a different group of units, and the results compared to outcomes in a set of control units. A synthetic control with two-treatment arms and one control arm could work well for this. A significant drawback of this approach is that it requires broad data collection on a granular scale. This is likely not feasible in most contexts.

Augment Existing Data with New Sources

If price data prove difficult to obtain, perhaps other data which are highly correlated with prices can serve to augment or proxy for price data.

Data Fusion

Data Fusion is a Bayesian technique arising from modelling data with missing values. For an overview see Feit and Bradlow (2021). Both in academic literature and in practice, **Data Fusion** is understudied and has major potential to help with data scarcity. Despite the wealth of data *in general*, data scarcity with respect to *specific problems* is still a major challenge. **Data Fusion** allows for different datasets to be used together in service of a modelling task. Bayesian modelling "fuses" datasets using common variables as a bridge. Common variables are those which both datasets contain. Suppose there are two datasets, A and B, with corresponding sets of variables V^A and V^B. Common variables are given by the set $V^A \cap V^B$. Denote by W_{bt} the matrix of variables, where each column is a common variable, and each row is an observation (at one brand, year combination).

It may not be immediately clear how Data Fusion can help with cash transfer data scarcity, especially when that data scarcity is in y_{bt}, as opposed to X_{bt} and Z_{bt}. The potential benefit comes from the ability to bridge multiple sources of data that reflect on y_{bt} using the common variables. Suppose we use y_{bt}^A to denote a measure of sales from one source (imagine market share data in a region containing the data scarce focal area), and in a region containing the data scarce focal area), and y_{bt}^B to denote a different measure of sales from a different source (perhaps sales at a local. A sketch of a typical process follows.

1. Build a Data Fusion model relating y_{bt} to y_{bt}^B and y_{bt}^A using common variables W_{bt} as a bridge. The specific model, $f(y_{bt} \mid y_{bt}^A, y_{bt}^B, W_{bt}, V^A, V^B)$, depends on the setup and assumptions.
2. Estimate model parameters (Δ) using the Data Fusion model.
3. The intended outcome is that the estimates from Data Fusion $\hat{\Delta}^{DF}$ are more accurate (as judged by the ability to predict the impact of AMT_{bt}) than the estimates using either source of data alone (say $\hat{\Delta}^A$ and $\hat{\Delta}^B$).

There are multiple Data Fusion setups that could potentially help with data scarcity, but we will focus on two. The first setup is to use cheap lab or field experiments as a source of clean, precise data, in addition to whatever observational sales data are available. The second setup is to use aggregate data moments (like regional market shares) to constrain model estimates in a way that improves accuracy. We also discuss **Surrogate Models,** which are similar to and could be considered Data Fusion, but have an independent literature, so we consider them in a separate section, below.

Conjoint and Other Experiments

Experiments can anchor informative priors or validate models, as described above, but they can also produce useful data for fusing with observational sales data. Conjoint analysis is a tool from marketing that uses a discrete choice experiment to elicit preferences across goods. Conjoint could be used as a way to infer market prices by presenting respondents with different consumption bundles, paired with a pure cash outside option. If the outside option were varied, the implicit prices of the bundles of goods would be identified. This would sidestep the need to collect actual in-market price data, and would instead tap into the latent knowledge about prices held by transfer recipients. Alternatively, a conjoint analysis could be used as an additional measure for variables affected by data scarcity, and this could be paired with observational data using Data Fusion. This is sometimes referred to as "data enrichment" (Kappe et al. 2017). The applications for this approach are extensive, because conjoint studies can be tailored to precisely generate whichever variables are desired. These variables could be predictors of types of expenditure, share of wallet, substitution across goods or expenditure categories, and much more. A challenge in these contexts is the gap between what cash transfer models typically include versus the controlled environment in the lab, along with the difficulty of testing too many factors in lab settings. However, lab studies can be used to add rich information about a specific variable of interest. Another option for primary (as in 1st party) studies is the survey. What follows is a sketch of a primary study paired with an observational model using Data Fusion.

1. Among a target recipient group defined by common variables W_{bt} (say, a set of demographics which are available in the data-scarce region), recruit a sample of 200 respondents.
2. In return for the cash, each recipient responds to a conjoint survey, and estimate implied market prices.
3. Use data fusion with data from enumerators and the conjoint data to predict outcomes or impute price data.

Aggregate Moments

The Generalized Method of Moments, frequently used in the econometrics literature, can be applied to data fusion problems, in the spirit of Petrin (2002): micro moments, generated from individual, store or community-level data, can be combined with meso/macro moments from regional observational data, to create estimates of parameters which are informed by both sets of data (for a recent example of this approach in a business context see McCarthy and Oblander, 2021). An alternative is to use Bayesian Moment Condition (BMC) Models as described in Chib et al. (2018). Such an application could take the same form as the primary data study, but more likely it would involve disaggregate data from some locations, along with aggregate data across a larger region. The two sets of moments would be included as conditions in the model, and parameters common between them estimated. Other than the moments specified, the estimators are non-parametric and so are less prone to misspecification bias. A sketch of an application follows.

1. Among a set of units with data, create moment conditions defining the model, say $E[y_{bt} - A - AMT_{bt}\Delta - X_{bt}\beta - Z_{bt}\delta] = 0$
2. For a set of aggregate moments, create a second set of moment conditions defining the same model, assuming the parameters are the same: $E\left[Y_{bt} - A - AMT_{bt}\Delta - X^A_{bt}\beta - Z^A_{bt}\delta\right] = 0$, where Y_{bt} is an aggregated outcome across the unit, and the superscript A also denote aggregated variables.
3. Use GMM or BMC models to estimate Δ, β, δ

Surrogate Models

Originally developed for causal inference over the long-term (Athey et al. 2019), surrogate methods have wide potential applicability. They could be used to model long-term impacts of cash transfer interventions, or they could be used to mitigate data-scarcity. The key idea of surrogate models is to find an easily accessible variable or set of variables that react to an intervention (e.g., cash transfer) which are highly correlated with an ultimate outcome of interest (say, long-term recipient welfare measures). A sketch of the method applies to data-scarce contexts is below:

1. Identify the surrogates (S_{bt}, e.g., digital trace data such as public social media posting or popular Google searches)
2. Train a model to predict outcomes in scarce areas (y_{bt}) using surrogates (S_{bt}). One could use whatever model works best. However, the model would need to be transparent enough to learn variable importance (tree ensembles, regularized regression, RuleFit, TabNet)
3. Estimate a short-term model but with the goal of maximizing outcomes as predicted by the surrogate index. Predictions of outcomes come from (2).
4. Create a policy recommendation from (3) and use it to run an experiment. The recommended testing regime is to use both a test and control cash transfer policy. The control policy uses available data and the test policy uses surrogates in addition to available data. Both policies produce transfer recommendations, and the causal effect of interest for the experiment is the difference in measurable long-term outcomes when using surrogate-informed vs. traditional data.

Google Search Data

Many of the above methods make use of auxiliary data. Google Trends is one possibility. However, there is little academic research validating this is a source of data, and Google has changed the measurement approach multiple times in the past. To be reliably used as a surrogate indicator, a validation test would need to be conducted. The validation would need to not only test the feasibility of using Google Trends data, but also identify *how* Google Trends should be used: which search terms among the enormous set of possibilities would be the best surrogate? To predict how demand for particular items might affect price changes, Google Trends may need to be paired with an estimate of market size, which could be tricky. However, it may be the case that when used as an augmenting variable or as a surrogate, Google Trends data *does* predict well. The evidence for or against this is relatively thin.

Reduce Negative Consequences of Scarce Data

Uncertainty Quantification

Bayesian Optimization is a technique that guides on-line data collection (often in experiment design or machine learning hyperparameter tuning) through the quantification of uncertainty about estimates along one of the dimensions in the data. The typical rule is that data are collected where uncertainty is largest. This helps to allocate data collection resources.

Three potential uses for uncertainty quantification are (a) targeting regions for applying data fusion, surrogates, or granular data collection (using third parties), (b) developing conditional bounds on estimates, and (c) targeting parameters for needed precision. For (a), any Bayesian method admits this automatically: simply find regions with the largest posterior predictive intervals. Machine Learning methods would follow an analogous approach. For (b), the posterior predictive distribution of a model from data-rich regions could be used to get predictive intervals for data-scarce regions. This would in turn generate a range of possible parameter values in the data-scarce regions. These upper and lower bounds may be useful for many purposes motivating typical cash transfer models. Finally, for (c), simply by measuring the posterior credible intervals of model parameters, parameters which lack precision are revealed. All of the above techniques can be applied to these uncertain parameters, using uncertainty quantification as a guide. It may be the case the estimates of some parameters are relatively precise, in which case effort can be channelled elsewhere.

A key feature of Bayesian methods that is likely to be particularly valuable when paired with uncertainty quantification is a managerially-defined loss function. The loss function takes as inputs "states of the world," which are different realizations of random variables. Loss functions are useful for thinking about the costs of using an inaccurate estimate. For example, suppose the goal of a cash transfer program is to provide as many households as possible with enough cash to buy a pre-determined basket of goods. Additionally suppose that the program designer faces uncertainty in the prices of the goods in the basket. If too little cash is given to each recipient, then there is a higher probability that the true basket price is higher than the transfer amount. On the other hand, as more cash is given to each recipient, the pool of recipients shrinks. In which direction should the policy err? This question is answerable with two ingredients: a measure of uncertainty, and a loss function. The loss function assigns a value to each outcome. The tools discussed above regarding elicitation of probabilities are applicable to loss function elicitation as well, and loss functions can be easily aggregated. With a loss function in hand, simply integrating the loss function over the posterior distribution of the model under different policies will product different policy-loss values, and the lowest loss policy can be selected.

Methods Conclusion

From the above menu of possible methods, an analyst could quite easily combine several together. For example: a **Bayesian Hierarchical Gaussian Process** model could be paired with **Data Enrichment** via **Informative Priors**. Or for a non-Bayesian example: **Google Trends** could be used in a **Surrogate** model, where the resulting parameter estimates were "shrunk" towards regional-level estimates using the **Empirical Bayes** formula above. It is difficult to give a specific recommendation without knowing the composition of a team, but hopefully the sketches are enough of a start that a group could generate a self-guided program of model development.

1.3. Next Steps

AI for Monitoring Prices

1. Large Language Models

This chapter explores how Large Language Models (LLMs) can be used to analyze text data and provide insights on price changes of essential goods in these areas.

Methodology

There are two components to this method, with substantial flexibility in each: data sources, and the selection of an LLM.

Data Sources

Ultimately, aid recipients are the best source of text data in terms of informational value. Unfortunately there is an obvious trade-off between information value and ease of acquisition. Some potential data sources that may strike a good balance between information value and ease of acquisition:

- Social media posts (e.g., X, Facebook, Reddit)
- News articles (from local and international organizations)

- Solicited text input from aid recipients (possibly via SMS)

Social media posts are likely to contain mostly second hand information, but they are widely accessible and updated in real-time. News articles may be slower but possibly vetted more thoroughly. Solicited text input has high potential as a data source, but could be very difficult to collect. SMS data collection can be implemented using an automated service (e.g., Twilio, Bird.com, CallHub) or text messages using WhatsApp Flows.

Large Language Model Selection

For ease of implementation, it would be best to avoid fine-tuning a large language model (LLM) and instead use an "off-the-shelf" solution. This means either an API or an open-weights model. In terms of general purpose LLM performance, LMSys maintains a leaderboard by ELO ratings. ELO ratings come from chess, and are computed from many pairwise preference ratings provided by human users. In the LMSys ratings, two LLMs answer the same prompt, and human users choose the one they prefer. ELO ratings attempt to account for the win rate of each LLM as well as the difficulty of the opponents (see Chiang et al., 2024 for a detailed description).

GPT-4-Turbo, Claude 3, Gemini, and Llama 3 all have high LMSys ELO rankings. However, to get an idea of running costs, Artificial Analysis maintains a similar leaderboard for quality, speed, and price. Llama 3 appears to be a very good choice for balancing cost, speed, and response quality. [1]

There are many ways to interact with LLMs, but a prominent Python library for running local LLMs (such as Llama 3) is LangChain, which serves as a useful interface between a terminal and an LLM (see here for the Quickstart guide).

Another option is EDSL, a new Python library developed by the startup Expected Parrot (see here for the Starter Tutorial). EDSL is designed specifically for extracting structured information from unstructured text, which could allow faster deployment.

Prompts and Output

This section discusses LLM output structure. An outline of the process for using an LLM would probably include three components:

1. Identify frequently mentioned goods (e.g., rice, wheat, cooking oil).
2. Assess trends related to these goods (e.g., rising prices, limited availability).
3. Generate summaries that highlight these trends (e.g., "People are reporting significant price increases for staple foods, particularly rice").

[1] Based on our current business partnership with META, owner of Llama, we may propose them to support this project

It may be the case that a single prompt can elicit all of 1-3, but it is not at all unlikely that multiple prompts would produce better results. EDSL is designed to extract this kind of data from unstructured text. It would likely require multiple rounds of prompting, but it could easily loop over a large corpus of texts and collect the results into a structured format like a Pandas DataFrame or a CSV file.

By design, EDSL is more likely to produce consistent structure over time, whereas LangChain is better for interactive use. If the intended output is a report or a table, EDSL might be the better choice. If the intended use is interactive consultation by analysts, LangChain might be better.

Extensions

Since LLMs are highly flexible, the resulting system of data collection followed by LLM synthesis could be put to many additional uses, such as:

- Identifying locations of greatest need.
- Monitoring the effect of interventions.
- Anticipating price dynamics.
- Understanding adaptation mechanisms.
- Learning more about the full chain of cause and effect from crisis leading to specific price changes
- Interactive data collection. Unlike other tools, LLMs could become a means of both analysing *and* collecting data.

This generality is an advantage of building a pipeline for analysing unstructured text data using LLMs.

Challenges and Limitations

There are downsides to LLMs, which must be acknowledged.

- The accuracy of the LLM output depends on the quality and representativeness of the text data. This is of course not specific to LLMs, but rather it is specific to the text data.
- Output variability. LLMs will not give the same answer even with the same prompt, unless the model's "temperature" is turned to zero, in which case the model becomes (very nearly) deterministic. However, setting temperature to zero has uncertain performance impacts, and may degrade quality (as judged by humans in context).
- The need for human oversight. LLMs may produce incorrect summaries or assessments, and human judgement is still required.

2. Pre-LLM Natural Language Processing Tools

This section discusses how “traditional” Natural Language Processing (NLP) tools can be used to analyse text data. Traditional here refers to pre-LLM. At a high level, the benefit of traditional tools is speed, scalability, constancy, and cost. However, these tools are significantly less general than LLMs, so any application would need to have a high degree of structure.

Methodology

Just as with LLMs, a source of text data is needed. The same sources apply here as in the LLM case. The recommended tool is spacy, a natural language processing library in Python.

Open Source NLP: spacy

Spacy is an open-source NLP library with good out-of-the-box performance. It is also easy to fine-tune spacy for custom tasks or datasets. Spacy offers a wide range of NLP functionalities, including NER and sentiment analysis.

- Named Entity Recognition (NER) can be used to identify locations and specific products within the text data. Named Entity Recognition (NER) acts like a digital highlighter for text data. It scans text and identifies specific categories of words and phrases, like locations (cities, countries), organizations, dates, or prices. This allows for automatic extraction of data from text sources, which means that structured datasets can be produced from unstructured text. This information could then be used to track price changes for essential goods in specific areas.
- Sentiment analysis can be used to gauge the degree of negative vs positive language. This information can be helpful in understanding the severity of the situation with respect to specific goods or locations.

Sentiment analysis can be done using the spacytextblob library in Python. NER is built into spacy, but to recognize custom entities, such as prices, a model would need to be trained using annotated text data.

Process for Training a New Entity

A sketch of the process to train a model to recognize a new entity follows.

1. Data Annotation with Prodigy:

Prodigy is a tool for efficiently annotating text data. A human expert would annotate text for training. This would involve the following:

- Prepare data: first, text data containing prices would be gathered. This would come from the same data sources that would eventually be analysed, such as news articles and social media posts.
- Create an annotation task: the "PRICE" entity label within Prodigy would be defined, and users would manually highlight specific price mentions in a sample of text data. This could be a dollar amount, a range, or a percentage change, as long as all annotators remain consistent with the definition.

2. Training, Validation, and Testing:

After producing annotated data, the spacy model could be trained. The steps are:

- Annotated data are divided into training, validation, and testing sets. The training set teaches the model, the validation set determines which model and model settings are selected for application, and the testing set measures final performance of the model.
- Train the model. Use spacy's training functions with the annotated data.
- Evaluate the model's performance on the validation and testing sets. The recommended performance metric for NER is the F-score.

See the Prodigy tutorial for documentation and instructions, and this NER video tutorial for detailed instructions on adding a new entity to an NER model.

Additional Considerations

Traditional NLP tools have some limitations.

- Relative to LLMs, NER and sentiment labelling are narrow and may fail to take full advantage of the information in text.
- Human annotation is challenging and it can be an arduous process to define annotation rules.

3. Combining LLMs and Traditional NLP Tools

It may be possible to integrate Large Language Models (LLMs) and traditional NLP techniques. This would take advantage of the fact that both approaches take the same text data as input. Possible arrangements include the following.

1. LLMs for interactive analyst support, with NER and sentiment tools for structured data reporting. The same source of text data would be fed into the NLP tools and the LLM system, and in fact an LLM could be given the NER report to summarize.
2. LLMs, NER/sentiment, and human experience to triangulate price information. In this case, LLM and NER/sentiment reporting would be given to a human analyst to make a final judgement. The correspondence between the LLM and traditional NLP reports would either increase confidence in the findings or lead to further exploration.

SECTION 2: DIGITAL IDENTIFICATION

2.1. Introduction

In an era marked by rapid technological advancements, humanitarian cash transfer programs stand at a crossroads, poised to harness new opportunities while navigating the complexities of an increasingly digital landscape. The integration of data analysis and Artificial Intelligence/Machine Learning (AI/ML) technologies offers a beacon of hope, promising to revolutionise the effectiveness and scope of these vital programs. This report delves into the innovative use of non-traditional digital trail data and AI/ML to bolster humanitarian cash transfers, emphasising the imperative to ensure these technological strides benefit the humanitarian ecosystem positively (Holm-Nielsen et al., 2023).

As we explore the potential of digital and AI/ML tools in surmounting the obstacles faced by contemporary cash transfer systems, we will delve into their roles in data analysis, identity verification, and decision-making processes. These technologies, when judiciously applied, have the power to transform data into actionable insights, streamline the identification of beneficiaries, and enhance the precision of aid delivery. However, this potential comes with an ethical imperative, particularly given the vulnerability of the populations these programs aim to serve.

The landscape of humanitarian aid is as diverse as it is challenging, with conflicts, natural disasters, and crises unfolding across the globe, each with its unique context and set of needs. Technology and new data methods offer a beacon of hope, allowing for context-specific targeting and adaptive strategies that can respond to the dynamic nature of global crises. Yet, the adoption of these technologies is not without challenges. Ensuring data accuracy, protecting privacy, and maintaining ethical standards are paramount, especially when dealing with the world's most vulnerable populations.

This report aims to provide a comprehensive analysis of the state-of-the-art in humanitarian cash transfers, highlighting the existing and emerging challenges in this field. By showcasing real-life applications and drawing on scientific references, we aim to offer a narrative that is both informative and accessible, providing valuable insights and recommendations to practitioners and policymakers dedicated to enhancing humanitarian efforts across the globe.

2.1. Exploration of Non-Traditional Digital Trail Data

Non-traditional digital trail data encompasses the diverse range of information generated by individuals as they interact with digital environments. Unlike conventional data sources like surveys or administrative records, this data includes social media activity, geolocation information, online search trends, and e-commerce behaviors, offering a rich, real-time snapshot of human behavior and circumstances.

For instance, during the 2015 Nepal earthquake, remote sensing data played a crucial role in rapid damage assessments, facilitating targeted and efficient response efforts (Lallemant et al., 2017). Similarly, in the aftermath of the Syrian conflict, geolocation data was pivotal in tracking population movements, enabling humanitarian agencies to better plan and implement their interventions (Hernandez & Roberts, 2020).

The potential of digital trail data extends beyond crisis response to the sphere of cash transfer programs. In Togo, an innovative approach was taken where mobile phone data was leveraged to identify and support individuals impacted by the COVID-19 pandemic. The Togolese government, utilising AI and ML techniques, analysed call detail records to distribute cash transfers to those most affected by the crisis, demonstrating a novel use of digital data for humanitarian aid (Aiken et al., 2022). In Zimbabwe, mobile cash transfers were utilsed to provide emergency food security support in drought-affected communities. This approach not only facilitated rapid and targeted assistance but also offered insights into the effectiveness of digital cash transfers in enhancing food security during crises (Tirivayi et al., 2016).

The use of digital trail data, however, necessitates a careful consideration of ethical issues. Ensuring privacy, maintaining anonymity, and securing informed consent are paramount to uphold the dignity and rights of individuals. It is essential to balance the benefits of data utilization with the potential risks of surveillance, ensuring that humanitarian efforts remain respectful and people-centric (Weitzberg et al., 2021).

By harnessing non-traditional digital trail data, humanitarian organizations can gain unprecedented insights into the needs and movements of affected populations, enabling them to deliver more effective and timely assistance. Yet, this must be done with a steadfast commitment to ethical standards, ensuring that the pursuit of innovation does not compromise the fundamental principles of humanitarian aid.

2.2. Application of AI/ML Techniques in Data Analysis

AI/ML can process and analyse vast amounts of unstructured, incomplete data, typical in crisis contexts, to extract actionable insights. These technologies can enhance data quality by identifying errors or inconsistencies in identity data, a crucial aspect given the importance of accurate beneficiary identification in cash transfer programs.

2.2.1 Improving Data Quality

The integrity of data in humanitarian contexts is paramount. AI techniques can enhance data quality by identifying and rectifying errors, duplicates, and inconsistencies in collected information. This is especially critical in the context of beneficiary identification, where the accuracy of data directly impacts the efficacy of aid distribution.

For instance, AI algorithms can be employed to analyze beneficiary data for inconsistencies, such as multiple entries for a single individual or incorrect information. By cross-referencing data points and identifying patterns, AI can highlight anomalies that may indicate data quality issues. The WFP has already leveraged such AI applications to refine its beneficiary databases, ensuring that aid is delivered to the intended recipients (Oxfam International, 2011). A significant contribution to data quality improvement comes from analyzing alternative data sources. The study by Zheng et al. (2022) discusses the design of allocation rules in cash transfer programs, highlighting the potential of AI in developing more nuanced and effective targeting strategies by utilizing diverse data inputs.

Leveraging AI to harness alternative data sources for enhancing data quality in humanitarian settings, starting point is exploting the potential of diverse and unconventional datasets to offer a more nuanced understanding of beneficiaries, which is pivotal for the efficacious distribution of aid.

Taking transactional data as an initial point of consideration, this type of information, like mobile money transactions, can provide deep insights into the economic behaviors and needs of individuals. By employing machine learning models such as clustering algorithms or anomaly detection, AI can segment beneficiaries based on their transactional behaviors or pinpoint outliers that might indicate data inconsistencies or potential fraud. A real-world application of this can be seen in the use of mobile money transaction patterns to identify and support economically vulnerable populations in East Africa, where such data have been instrumental in tailoring financial assistance programs.

Shifting focus to communication patterns, the analysis of call detail records (CDRs) can yield indirect indicators of an individual's social and economic status. Network analysis applied to these CDRs can reveal the social capital of individuals, potentially indicating their vulnerability. For example, during the Ebola outbreak in West Africa, CDR analysis was utilised to understand population movements and social interactions, aiding in the containment strategies and resource allocation.

Digital footprints, such as social media activity or app usage patterns, are another goldmine for insights. By applying sentiment analysis or topic modeling on social media content, humanitarian organizations can gauge the concerns or needs of populations. This was notably demonstrated during the Syrian refugee crisis, where social media analysis provided insights into the needs and sentiments of displaced populations, informing humanitarian response strategies.

Geolocation data also plays a critical role, especially in disaster response or monitoring refugee movements. Time-series analysis or predictive modeling applied to this data can forecast movement patterns or identify areas with increasing concentrations of vulnerable individuals. For instance, geolocation data from mobile phones was crucial in tracking population movements after the 2015 Nepal earthquake, aiding in the coordination of relief efforts.

Integrating these diverse data sources through data fusion techniques offers a holistic view of the situation, enhancing the robustness and accuracy of the insights derived. However, it's imperative that the use of such data adheres to strict ethical standards, ensuring the privacy and consent of the individuals involved. The balance between innovation and ethical responsibility remains paramount as humanitarian organizations leverage these advanced technologies to improve their interventions and aid distribution.

2.2.2. Enhancing Identity Verification

In regions where robust foundational identity systems are lacking, AI offers groundbreaking methods for verifying individual identities. By tapping into non-traditional data sources, such as biometric data and behavioural patterns, AI facilitates the creation of dependable identity registries. While the WFP has initiated steps towards integrating AI and biometric technologies in their operations, exploring external innovative practices can further augment their identity verification processes.

The utilisation of AI in identity verification outside WFP offers insightful precedents. For instance, the application of AI in Togo's mobile phone data analysis to identify individuals facing economic challenges demonstrates an innovative approach to leveraging alternative data sources for identity verification in cash transfer programs (Aiken et al., 2022). Similarly, in Zimbabwe, AI-enhanced mobile cash transfers have been instrumental in ensuring that emergency aid reaches the intended beneficiaries, showcasing how AI can improve identity verification processes (Tesfaye, 2017).

A particularly relevant example for WFP to consider is the Aadhaar system in India, one of the world's largest biometric ID systems. The system uses biometric data for identity verification, significantly reducing fraud and improving the efficiency of government subsidy programs (Dhawan, 2020). By examining Aadhaar's architecture and implementation, WFP can glean valuable insights into scaling biometric systems, addressing privacy concerns, and enhancing user authentication.

Another example is the use of blockchain technology in conjunction with AI for identity verification, as explored in the humanitarian sector by the ID2020 alliance. This technology offers a secure, immutable ledger for storing biometric data, providing a robust framework for verifying identities while maintaining data privacy (Pisa & Juden, 2017).

2.3. Enhancing Digital ID Verification Through AI in Challenging Environments

2.3.1. Technical Suggestions for Improving Data Quality

Enhancing the quality of digital IDs is crucial, particularly in regions where traditional data sources are scarce or of poor quality. Leveraging Artificial Intelligence (AI) can significantly improve the reliability and accuracy of these digital systems through sophisticated techniques designed to meet the demands of challenging environments.

Advanced Image Processing Techniques: Employing machine learning algorithms such as convolutional neural networks (CNNs) is pivotal for enhancing image classification, recognition, and segmentation. These algorithms excel in managing complex image data captured under diverse environmental conditions, such as inconsistent lighting, varied angles, and low-resolution images. By training these models on extensive, diverse datasets, they become better equipped to handle real-world variations. For instance, CNNs have been effectively utilized in biometric identification systems in transient settings like refugee camps, where traditional ID mechanisms falter. The adaptive nature of CNNs allows for continuous learning and improvement, making them ideal for dynamic field applications where conventional data processing methods might fail (Bronstein et al., 2019).

Synthetic Data Generation: In areas with a dearth of reliable data, synthetic data generation emerges as a powerful tool. Techniques such as Generative Adversarial Networks (GANs) play a crucial role by producing a broad spectrum of realistic images that mimic a variety of real-world conditions. These images are used to train AI models, enhancing their ability to handle unexpected variations and increasing the robustness of the systems. For example, in facial recognition technology, GANs have been instrumental in creating diverse facial datasets that accurately represent different demographics, thus mitigating common biases in AI models and enhancing their applicability across various global contexts (Zhao et al., 2020).

Data Augmentation: To enhance the generalizability of AI models, data augmentation techniques are employed. These methods involve modifying images through various transformations such as rotation, scaling, cropping, and color adjustment. This diversification allows models to learn from a broadened dataset and perform more reliably under different operational conditions. Data augmentation is especially beneficial in fields such as medical imaging, where it helps to train algorithms on a wider range of scenarios, significantly improving diagnostic accuracy and reliability. Such techniques ensure that digital ID systems are not only adaptable to varied environments but also resilient to the variations inherent in global humanitarian settings (Wang et al., 2021).

By integrating these advanced techniques, WFP can create digital ID systems that are not only more accurate and efficient but also fair and inclusive. These systems are capable of performing reliably in diverse and adverse conditions, ensuring that aid and resources are allocated correctly and equitably. The strategic application of these AI-driven methods enhances the functionality and integrity of digital IDs, ultimately supporting more effective and responsive humanitarian operations.

2.3.2. Addressing Common Biases in Photo Identifications

Photo identification systems, integral to modern digital ID verification, often exhibit biases that can substantially undermine their accuracy and fairness. These biases, which span racial, gender, age distinctions, and environmental conditions, have significant ramifications for the equity of AI applications.

a. Intersectionality – Gender and Racial Bias: In humanitarian aid settings where access to critical services often hinges on accurate photo identification, the intersection of gender and racial biases in AI systems can have profound implications. These biases exacerbate vulnerabilities among the most marginalized groups, particularly women of colour, by potentially denying them access to essential aid due to inaccuracies in AI-driven identification processes. AI systems frequently exhibit gender biases, showing a higher efficacy in identifying males over females. This bias predominantly arises from datasets overwhelmingly composed of male subjects. The landmark "Gender Shades" study by Buolamwini and Gebru (2018) highlighted these discrepancies, particularly underscoring the lower accuracy rates in identifying women, and even more so for women of colour, stressing the urgent need for gender-diverse training datasets. Racial bias in AI manifests through the differential performance of photo identification systems across racial lines. The "Gender Shades" project also illuminated this issue, showing that commercial AI systems often have higher misidentification rates for individuals with darker skin tones, particularly dark-skinned women, due to a lack of diverse racial representation in training datasets.

Intersectionality refers to the combined effects of race and gender (and other identities) that compound discrimination or privilege. In humanitarian scenarios, such intersectional biases can critically hinder access to necessary aid for specific groups, most notably women of colour.

- **Heightened Risk in Aid Distribution:** In refugee camps or disaster relief operations where identity verification through photo ID is crucial for distributing aid, biases in AI can lead to the misidentification of beneficiaries. For example, dark-skinned women might be disproportionately affected, potentially leading to a denial of services like healthcare, shelter, and food supplies.
- **Real-World Consequences:** Inaccurate AI systems in these contexts not only impact individual lives but also undermine the efficiency and equity of aid distribution efforts. For instance, during the aid response to the 2010 Haiti earthquake, issues with biometric identifications were reported where racial and gender biases in technology hindered effective aid delivery to certain groups.

Strategies to Mitigate Intersectional Biases in AI for Humanitarian Aid:

- **Inclusive Data Collection and Training**: It's crucial to incorporate a broad spectrum of demographic groups in AI training sets, especially representing those frequently underserved or underrepresented. Including more images of women, various racial groups, and particularly individuals from these groups within humanitarian contexts is essential.

- **Algorithmic Transparency and Rigorous Testing:** Implementing transparent development processes and conducting extensive testing across diverse groups can help identify and address biases. AI systems used in aid distribution should undergo field testing to ensure they operate fairly across all demographics.
- **Adaptive AI Development:** Developing AI systems that can adapt to the unique challenges of humanitarian settings, including varied environmental and cultural contexts, is critical. These systems should be designed to continuously learn and improve their accuracy in real-time deployment.

b. Age Bias: Age bias in photo identification systems emerges as a significant challenge in humanitarian settings, where accurate identification is crucial for accessing vital aid and services. This bias occurs because AI systems often struggle to accurately recognize individuals across different age groups, particularly children and the elderly, due to the underrepresentation of these age demographics in training datasets. Enhancing the photo quality and updating systems in these vulnerable areas are essential steps towards equitable aid distribution.

In-depth Analysis and Real Examples:

Research from Levi & Hassner (2015), published in the IEEE Transactions on Information Forensics and Security, demonstrates that facial recognition technologies frequently misidentify younger and older age groups. The study underlines the critical need for training data that encompasses a broader age spectrum to improve the accuracy of these systems across all age groups.

- **Challenges for the Elderly and Children:** In humanitarian crises, such as in refugee camps or disaster-hit regions, the elderly and children are often the most vulnerable. AI systems that fail to identify these groups accurately can lead to serious consequences, including the misallocation of essential resources like food, medical aid, and shelter.
- **Real-World Example in Emergency Responses:** During the emergency response to natural disasters, such as the Nepal earthquake in 2015, aid organizations used biometric systems for distributing aid. Reports indicated that age bias in these systems resulted in inefficient aid distribution, where elderly individuals were less likely to be correctly identified due to the aged appearance changes not being reflected in the AI models used.

Strategies to Mitigate Age Bias in Humanitarian Settings:

- **Diverse Data Collection:** Ensuring that AI training datasets include a balanced representation of age groups is crucial. This can involve gathering images from various humanitarian contexts to reflect the real-world diversity of age groups encountered in these settings.

- **Continuous System Updates:** Regularly updating AI systems with new data that includes age-progressed images can help mitigate age bias. For sustained aid in long-term humanitarian crises, periodic re-enrolment or update of beneficiaries' photos in the system databases can maintain the accuracy of age identification.
- **Specialised Algorithm Development:** Developing algorithms specifically designed to detect and correctly identify age-specific features can improve recognition accuracy. Machine learning techniques that focus on adaptive learning from new inputs can continually refine the models based on the changing demographics within aid settings.
- **Inclusive Testing and Validation:** Conducting field tests of facial recognition technologies in diverse humanitarian environments ensures that the systems are robust and reliable across all age groups. This should include stress-testing the systems under various environmental conditions that mimic those found in aid distribution areas.
- **Community Engagement and Feedback:** Engaging with local communities to understand their specific needs and receiving feedback on the technology's performance can guide more targeted improvements. This approach ensures that the systems are not only technically sound but also culturally sensitive and practically useful.

c. Bias Due to Lighting and Environmental Conditions: Environmental factors, particularly lighting conditions, critically influence the performance of photo identification systems, and these can disproportionately affect individuals with darker skin tones. In humanitarian settings where access to aid is often linked to identity verification, the reliability of photo ID systems under diverse environmental conditions becomes crucial. Such disparities are not merely technical shortcomings but have profound implications for fairness and equity in the deployment of AI technologies.

In-depth Analysis and Real Examples:

- **Impact on Darker Skin Tones:** The research by Klare et al. (2012), published in the IEEE Transactions on Pattern Analysis and Machine Intelligence, highlights that facial recognition systems generally underperform in low-light conditions, particularly affecting individuals with darker skin. This underperformance can lead to higher misidentification rates in settings such as refugee camps or disaster-affected areas where lighting cannot always be controlled, and many residents may have darker skin tones.
- **Case Studies in Humanitarian Aid Distribution:** Inadequate photo ID systems can severely disrupt the distribution of aid. For example, during nighttime or in poorly lit temporary shelters, essential aid services relying on facial recognition for the distribution of food and medical supplies may fail to accurately identify individuals, leading to potential deprivation of necessary aid.
- **Case Study – Refugee Camps:** In refugee camps, where people from diverse backgrounds converge, the variability in environmental lighting can be significant. A study conducted by the

United Nations High Commissioner for Refugees (UNHCR) observed that the efficacy of biometric systems often decreases in outdoor setups, especially under variable weather conditions, impacting the accurate identification of refugees and thereby their access to regular assistance (UNHCR, 2019).

Steps to Mitigate Lighting and Environmental Conditions Bias:

- **Advanced Imaging Technologies:** Implementing technologies such as infrared and thermal imaging can significantly improve the robustness of cameras in low-light environments. These technologies are crucial in humanitarian settings where traditional lighting solutions may not be feasible.
- **Algorithmic Improvement and Training:** Enhancing algorithms to handle variations in lighting is vital. This includes training AI models with datasets that represent a wide range of skin tones under varied lighting conditions. A notable initiative by MIT and Stanford researchers has demonstrated that machine learning models trained on diverse datasets exhibit fewer biases and higher accuracy across different demographics (Buolamwini & Gebru, 2018).
- **Adaptive Systems for Environmental Variation:** Developing systems that adapt to changes in lighting by automatically adjusting camera settings or processing techniques can mitigate bias. Techniques such as dynamic range adjustment and exposure compensation should be integrated to enhance image quality in real-time, ensuring more reliable identification.
- **Rigorous Testing and Validation:** Regular testing of facial recognition systems under different environmental conditions is essential. This should include field tests that simulate the challenging settings often found in humanitarian aid scenarios to better understand how these systems perform outside controlled laboratory conditions.

d. Cultural and Religious Attire Bias: Cultural and religious attire biases in photo identification systems represent a significant challenge in the deployment of AI technologies. These systems often struggle to accurately identify individuals who wear garments such as hijabs, turbans, kippahs, or other headgear, which are integral to their cultural and religious identities. The root cause of these inaccuracies typically lies in the underrepresentation of these groups in the training datasets used to develop AI models.

In-depth Analysis and Real Examples:

- **Challenges for Muslim Women:** One of the most documented instances of this bias pertains to Muslim women wearing hijabs. The Georgetown Law Center on Privacy & Technology highlighted that facial recognition technologies frequently misidentify Muslim women, potentially leading to significant consequences in security, legal, and social contexts (Garvie et al., 2016). These misidentifications stem from AI's inability to process the unique facial visibility in hijab-wearing women adequately, as traditional datasets predominantly include uncovered heads and faces.

- **Sikh Community Issues:** Similarly, individuals wearing turbans, such as those from the Sikh community, face recognition challenges. A study conducted by the AI Now Institute pointed out that standard facial recognition systems often fail to distinguish between different individuals wearing similar style turbans, as the covering interferes with the algorithms' ability to map facial features accurately (Raji & Buolamwini, 2019). This can lead to a higher rate of false positives or negatives, which is particularly problematic in law enforcement and security screenings.
- **Impact on Jewish Practices:** The Jewish practice of wearing kippahs and other traditional head coverings also poses identification challenges. Research from the Massachusetts Institute of Technology (MIT) has shown that facial recognition software has a harder time processing images when customary head coverings obscure parts of the hairline and ears, which are critical landmarks for these algorithms (Buolamwini & Gebru, 2018).

Steps to Mitigate Cultural and Religious Attire Bias:

- **Diversified Data Collection:** To counter these biases, it's crucial to incorporate a more diverse range of images in training datasets that include individuals wearing various cultural and religious attire. This approach helps the AI learn to recognize and process a broader spectrum of human appearances.
- **Community Engagement:** Engaging with communities to understand the specific challenges and nuances of cultural attire can inform the development of more accurate and respectful AI systems. This includes direct involvement from community representatives in the design and testing phases of system development.
- **Algorithmic Adjustments:** Developers can employ advanced algorithmic techniques that focus on parts of the face not obscured by attire or improve the AI's ability to use partial facial recognition when certain features are covered. Techniques like neural architecture search (NAS) can be used to develop customized models that better handle the variability introduced by different attire.
- **Ethical Standards and Auditing:** Implementing regular audits and adhering to ethical AI development standards can help identify and correct biases proactively. These audits should be conducted by independent third parties to ensure objectivity and include feedback from affected communities to make necessary adjustments.

e. Disability Bias: Disability bias in photo identification systems arises when AI technologies inaccurately identify individuals with physical disabilities affecting facial features. This bias is often rooted in the training datasets that largely comprise images of 'typical' symmetrical facial features, neglecting the variations presented by conditions like Bell's palsy or muscular dystrophy.

The repercussions of such oversights are particularly acute in humanitarian contexts where access to essential aid is contingent upon reliable identification.

In-depth Analysis and Real Examples:

- **Challenges for Individuals with Facial Paralysis:** Conditions such as Bell's palsy can cause temporary or permanent asymmetries in facial features, posing significant challenges for standard facial recognition algorithms. A study documented in the International Journal of Technology Assessment in Health Care by Hamm et al. (2021) illustrates that facial recognition systems frequently misidentify individuals with such conditions, attributing to the systems' heavy reliance on facial symmetry for identity verification.
- **Misidentification Issues in Humanitarian Aid Distribution:** In humanitarian aid contexts, misidentification can prevent individuals with facial disabilities from accessing critical services and support. For example, in refugee camps where aid distribution often relies on biometric systems, those with conditions that alter facial muscle tone, such as muscular dystrophy, may find themselves repeatedly misidentified, leading to aid disruption. These issues underline the need for systems that can adapt to diverse facial presentations to ensure equitable access to assistance.
- **Legal and Ethical Concerns:** The Equality Act 2010 in the UK mandates reasonable adjustments to ensure that services are accessible to everyone, including those with disabilities. Failure of facial recognition systems to accommodate individuals with disabilities could lead to legal challenges based on discrimination, underscoring the necessity for compliance with legal standards in AI deployment (Equality and Human Rights Commission, 2020).

Steps to Mitigate Disability Bias:

- **Diversified Data Collection:** Enhancing the diversity within facial recognition datasets is crucial. Collaborating with organisations that advocate for individuals with disabilities can help collect a broader array of images that encompass a wide spectrum of facial differences. This approach ensures that AI systems are trained on data that reflect the true diversity of human features.
- **Algorithmic Adjustments:** Improving algorithms to better recognise and process non-symmetrical facial features is vital. Employing machine learning models trained on asymmetrical data can significantly enhance the accuracy of facial recognition systems and reduce misidentification rates. Such technical refinements should be a priority in the development of AI systems used in critical applications like humanitarian aid.
- **Inclusive Design and Testing:** Implementing an inclusive design philosophy from the initial stages of AI development ensures the integration of diverse needs. Involving individuals with disabilities in the design and testing phases allows for real-time feedback, which is crucial for

refining systems to better serve their needs. This participatory approach not only improves the technology but also aligns with ethical standards by actively including affected communities in the development process.

- **Regular Auditing and Compliance:** Establishing regular auditing processes to assess the performance of facial recognition systems in identifying people with disabilities is essential. These audits should verify compliance with legal standards and help identify any persistent biases or inaccuracies, prompting timely adjustments.

f. Ageing Bias: Ageing bias in photo identification systems represents a significant challenge, particularly in humanitarian contexts where consistent access to aid depends on reliable identification. As individuals age, their facial features naturally change, but if AI systems are not regularly updated with recent images, their effectiveness diminishes over time. This is crucial in vulnerable areas, where photo identification is frequently used to distribute aid, and the repercussions of misidentification can lead to deprived access to essential services and support.

In-depth Analysis and Real Examples:

- **Impact on Long-term Aid Recipients:** In settings where humanitarian aid is distributed over extended periods, such as refugee camps or disaster recovery zones, recipients often experience changes in physical appearance due to ageing, environmental conditions, and stress. A longitudinal study by Best-Rowden et al. (IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018) demonstrated that without regular updates, the accuracy of facial recognition systems decreases, which can be particularly problematic in these contexts where consistent identity verification is crucial for ongoing aid distribution.
- **Challenges in Resource-Limited Settings:** The difficulty of updating photographic databases is exacerbated in resource-limited settings typical of humanitarian aid environments. For instance, in regions affected by conflict or natural disasters, technology infrastructure may be compromised, and capturing high-quality, updated photographs may not be feasible without significant logistical efforts.
- **Case Study - Aid Distribution in Conflict Zones:** In conflict zones like Syria or Yemen, where individuals may remain in refugee camps for many years, ageing bias can prevent the accurate identification of long-term residents, complicating their access to regular food supplies, medical aid, and resettlement programs. Organizations operating in these areas have observed that older photos fail to match current appearances, leading to delays and errors in aid distribution.

Steps to Mitigate Ageing Bias:

- **Regular Image Updates:** Implementing a system for regular photographic updates is essential. This could involve setting up periodic photo stations within aid distribution centers or using mobile devices to capture and update photos during regular aid distribution rounds.

- **Use of Age-Progression Technology:** Integrating age-progression technology in photo identification systems can help predict and adjust for changes in individuals' appearances. This technology, initially developed for locating missing persons, can be adapted to predict how a person's facial features might change over time, enhancing the system's ability to accurately identify individuals as they age.
- **Training on Diverse Age Groups:** Enhancing AI models by training them on a broad spectrum of age groups, especially focusing on the ageing process under various environmental stresses typical of humanitarian settings, can improve recognition accuracy. This training must include data from individuals who have aged under similar conditions to those experienced in humanitarian crises.
- **Community-Based Feedback Systems:** Establishing feedback mechanisms within communities to report and rectify identification errors can help maintain the accuracy of photo identification systems. Community involvement ensures that the systems are responsive to the specific needs and challenges faced by the population.

2.4. Structured Approach to Localisation and Regional Training for Photo Identification Systems

In humanitarian aid and crisis settings, the deployment of photo identification systems that accurately reflect and serve diverse populations is critical. Localisation and regional training form the backbone of this approach, enabling technology to adapt to the unique demographic, cultural, and environmental variations specific to each region. This tailored approach is essential due to the substantial differences in facial features, expressions, and attire across global populations, which generic models often inadequately capture.

Localisation in photo identification involves adjusting AI models to accurately recognise and interpret the varied characteristics of different regional populations. This process is not only a technical necessity but also a matter of ethical urgency, ensuring that all individuals receive equitable access to aid and services without bias or misidentification.

Technical Solutions for Localisation:

- **Fine-Tuning Pre-Trained Models:** Fine-tuning involves making small adjustments to a pre-trained model to adapt it to local specifics. This process requires fewer resources than training a model from scratch and leverages the learned features of the global model while adapting it to recognize local nuances. For instance, a model trained on a diverse global dataset can be fine-tuned with a smaller, locally relevant dataset that includes specific ethnic features, attire, or environmental backgrounds characteristic of the target region.
- **Transfer Learning:** Transfer learning is a technique where a model developed for one task is reused as the starting point for a model on a second task. This is particularly useful in adapting models to new regional contexts by transferring knowledge from general settings to specific

local conditions. An AI model trained to recognise faces in a European context could be adapted to an African context by retraining the model's final layers with images that represent the local demographic, ensuring it accurately captures the diverse features of the population.

- **Domain Adaptation:** Domain adaptation techniques are used to adjust AI models to new, unseen environments, helping overcome issues like varying lighting conditions or backgrounds that differ significantly from the training data. Adapting a facial recognition system to work in the arid landscapes and intense sunlight of refugee camps in sub-Saharan Africa, for example, would involve adjusting the model to perform well under these specific lighting and background conditions.
- **Incremental Learning:** Incremental learning allows AI systems to continually learn from new data, updating themselves to reflect changes in the environment or population without needing a complete retraining. In a region experiencing rapid demographic changes due to migration or conflict, an incremental learning approach allows the photo ID system to adapt continuously as new types of facial data are collected, ensuring the system remains effective and up-to-date.

The Imperative for Localisation in AI Systems:

Localisation of AI technologies is vital to ensure that photo identification systems accurately reflect and serve the diverse characteristics of various populations. Tailoring AI models to regional specificities involves technical adjustments that account for local demographic, cultural, and environmental variations, which are essential for the systems' accuracy and acceptability in different contexts.

Challenges and Technical Solutions for Localisation:

- **Adapting to Regional Variations:**
  - <u>Technical Approach:</u> Implementing fine-tuning and transfer learning techniques allows pre-trained global models to adapt to local features, such as ethnicity-specific facial characteristics, local attire, and even environmental factors like lighting conditions typical of certain regions.
  - <u>Real-World Application:</u> In regions like sub-Saharan Africa, where diverse ethnic groups present unique identification challenges, AI models can be refined to better recognise and process these variations, enhancing the accuracy of aid distribution.
- **Continuous Model Updating:**
  - <u>Technical Approach:</u> Utilising incremental learning and domain adaptation strategies ensures that AI systems evolve in response to new data and changing environments, maintaining their relevance and effectiveness over time.

    - Real-World Application: In conflict zones or areas with high migration flows, AI systems must frequently update to reflect the changing demographics, ensuring that identification remains reliable and effective.
- **Ensuring Ethical Compliance and Cultural Sensitivity:**
    - Technical Approach: Developing robust frameworks for ethical AI use that incorporate privacy, data protection, and bias mitigation is essential. This includes setting up Ethical Boards to oversee AI deployments and ensure alignment with humanitarian principles and local cultural norms.
    - Real-World Application: Establishing an Ethical Board to regularly assess the AI systems used in Middle Eastern refugee camps could ensure that the technology respects cultural sensitivities and adheres to ethical standards.

2.5. Conclusion

The strategic integration of AI ans ML into humanitarian operations, specifically for improving the data quality of photo identification systems, represents a critical advancement in the field of humanitarian aid. This report has elucidated the transformative potential of AI and ML technologies in refining the accuracy and efficiency of identity verification processes, which are fundamental in crisis-hit regions around the globe. For the WFP, the path forward involves leveraging these technological innovations to enhance the delivery of aid, setting a benchmark for the humanitarian sector in technological adoption that aligns meticulously with the organization's core humanitarian mission.

Strategic Imperatives for Enhanced Photo Identification in Humanitarian Contexts:

- **Robust Data Quality and Verification Systems:** Utilising AI to enhance the quality and reliability of data in photo identification systems is essential. AI algorithms can help identify and correct inconsistencies in data entries, verify identities with greater accuracy, and reduce fraudulent activities. For example, incorporating machine learning models to cross-validate beneficiary information across various databases can significantly improve the accuracy of aid distribution, ensuring that resources reach the rightful recipients in crisis zones.
- **Localisation of AI Models:** Adapting AI models to recognize and effectively interpret the diverse characteristics of different populations in crisis-affected areas is crucial. This involves using transfer learning and domain adaptation to tailor AI tools to local demographic and environmental conditions, which vary significantly across regions. By customizing these models, WFP can ensure that the identification systems are sensitive to local nuances, thereby increasing their effectiveness and acceptance within local communities.
- **Ethical Implementation and Oversight:** As WFP integrates AI into its operations, establishing an Ethical Board to oversee the deployment of these technologies is imperative. This board

should ensure that all AI applications adhere to the highest standards of ethical practice, including considerations for privacy, data protection, and fairness. This oversight will help mitigate potential biases and ethical pitfalls, maintaining the dignity and respect of individuals in vulnerable populations.

- **Building Technological Capacity and Inclusive Engagement:** Enhancing the technological proficiency of WFP staff and local partners is essential for the successful implementation of AI and ML technologies. Developing targeted training programs and continuous learning opportunities will equip them with the skills needed to manage these advanced tools. Moreover, engaging with local communities to gather feedback and refine the technologies will ensure that the solutions developed are well-suited to the users' needs and conditions.
- **Collaboration and Innovation:** Foster collaborations with technology leaders, academic institutions, and other humanitarian organizations to explore new AI applications that can be adapted for humanitarian use. These partnerships can lead to innovative solutions that address specific challenges faced in crisis settings, such as scalable and adaptable photo identification systems designed for dynamic and complex environments.

A Vision for the Future:

Embracing AI/ML in photo identification processes offers WFP a formidable opportunity to redefine the standards of aid delivery in humanitarian settings, particularly in regions experiencing long-standing or acute crises. By implementing the strategies outlined in this report, WFP can enhance its operational capabilities, delivering aid more effectively and with greater accountability.

This commitment to technological advancement, paired with a steadfast focus on ethical practices and community engagement, will position WFP as a leader in the humanitarian sector, advocating for and demonstrating the profound impact that responsibly applied technology can have on global humanitarian efforts. The integration of these advanced tools into WFP's workflows is not merely a functional enhancement but a strategic investment in a future where technology and humanitarian aid are seamlessly intertwined, leading to improved outcomes for those in critical need.

References

References for Section 1

Aiken, E., Bellue, S., Karlan, D., Udry, C., & Blumenstock, J. E. (2022). Machine learning and phone data can improve targeting of humanitarian aid. *Nature*, 603(7903), 864–870. https://doi.org/10.1038/s41586-022-04484-9

Albu, M. (2010). *The emergency market mapping and analysis toolkit*. Practical Action Pub.

Barrett, C. B., Bell, R., Lentz, E. C., & Maxwell, D. G. (2009). Market information and food insecurity response analysis. *Food Security*, 1(2), 151–168. https://doi.org/10.1007/s12571-009-0021-3

CALP Network. (2021).Data responsibility toolkit: a guide for CVA practitioners (report).

Cavallo, A., Cavallo, E., & Rigobon, R. (2014). Prices and supply disruptions during natural Disasters. *Review of Income and Wealth*, 60(S2). https://doi.org/10.1111/roiw.12141

Cumbane, S. P. (2022). P*opulation displacement estimation during disasters using mobile phone data* (thesis). KTH School of Architecture, Stockholm.

Esmalian, A., Coleman, N., Yuan, F., Xiao, X., & Mostafavi, A. (2022). Characterizing equitable access to grocery stores during disasters using location-based data. *Scientific Reports*, 12(1). https://doi.org/10.1038/s41598-022-23532-y

Idris, I. (2016). *Economic impacts of humanitarian aid* (report). Applied Knowledge Services.

Informal Businesses. World Bank. (n.d.). https://www.enterprisesurveys.org/en/informal-businesses/methodology

Informal economy: data futures exchange. UNDP. (n.d.). https://data.undp.org/insights/informal-economy

Inter Agency Standing Committee. (2023). *Operational Guidance on Data Responsibility in Humanitarian Action* (report).

Gassmann, F., Gentilini, U., Morais, J., Nunnenmacher, C., Okamura, Y., Bordon, G., & Valleriani, G. (2023). *Is the magic happening? a systematic literature review of the economic multiplier of cash transfers*. The World Bank. https://doi.org/10.1596/1813-9450-10529

Henein, S. (2023). *Natural disasters and agricultural commodity prices: global evidence* (thesis). Concordia University, Montreal.

Huddleston, R. J., & Wood, D. (2021). Functional markets in Yemen's war economy. *Journal of Illicit Economies and Development*, 2(2), 204. https://doi.org/10.31389/jied.71

Idris, I. (n.d.). Economic impacts of humanitarian aid.

Rapid assessment for markets: guidelines. (2014) . International Committee of the Red Cross.

Jung, W. (2023). Mapping community development aid: spatial analysis in Myanmar. *World Development*, 164, 106124. https://doi.org/10.1016/j.worlddev.2022.106124

Kamara, N. (2013). *Comparative review of market assessments methods, tools, approaches and findings*. World Food Programme.

Kuner, C., & Marelli, M. (2020). *Handbook on data protection in humanitarian action*.

Ragini, J. R., Anand, P. M. R., & Bhaskar, V. (2018). Big data analytics for disaster response and recovery through sentiment analysis. *International Journal of Information Management*, 42, 13–24. https://doi.org/10.1016/j.ijinfomgt.2018.05.004

Rahwan, I. (2018). Society-in-the-Loop: programming the algorithmic social contract. *Ethics and Information Technology*, 20(1), 5–14. https://doi.org/10.1007/s10676-017-9430-8

UNDRR.(2023). *UNDRR Data Strategy and Roadmap 2023–2027*. United Nations Office for Disaster Risk Reduction.

UNDP.(2023). *Digital Strategy 2022-2025*. United Nations Development Programme.

UN Secretary-General. (2020). *Data strategy of the secretary-general for action by everyone, everywhere*.

References for Section 1- Technical Methodology Section

Abadie, Alberto. "Using synthetic controls: Feasibility, data requirements, and methodological aspects." *Journal of Economic Literature* 59.2 (2021): 391-425.

Ashwin, Julian, Aditya Chhabra, and Vijayendra Rao. "Using Large Language Models for Qualitative Analysis can Introduce Serious Bias." *arXiv preprint arXiv:2309.17147* (2023).

Athey, Susan, Raj Chetty, Guido W. Imbens, and Hyunseung Kang. *The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely*. No. w26463. National Bureau of Economic Research, 2019.

Chiang, Wei-Lin, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Hao Zhang et al. "Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference." *arXiv preprint arXiv:2403.04132* (2024).

Chib, Siddhartha, Minchul Shin, and Anna Simoni. "Bayesian estimation and comparison of moment condition models." *Journal of the American Statistical Association* 113, no. 524 (2018): 1656-1668.

Chopra, Felix, and Ingar Haaland. "Conducting qualitative interviews with AI." (2023).

Dew, Ryan, Asim Ansari, and Yang Li. "Modeling dynamic heterogeneity using Gaussian processes." *Journal of Marketing Research* 57, no. 1 (2020): 55-77.

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker. "General equilibrium effects of cash transfers: experimental evidence from Kenya." *Econometrica* 90, no. 6 (2022): 2603-2643.

Feit, Elea McDonnell, and Eric T. Bradlow. "Fusion modeling." *Handbook of Market Research*. Cham: Springer International Publishing, 2021. 147-180.

Gelfand, Alan E., and Erin M. Schliep. "Spatial statistics and Gaussian processes: A beautiful marriage." *Spatial Statistics* 18 (2016): 86-104.

Gelman, Andrew, and Cosma Rohilla Shalizi. "Philosophy and the practice of Bayesian statistics." *British Journal of Mathematical and Statistical Psychology* 66, no. 1 (2013): 8-38.

Kappe, Eelco, Sriram Venkataraman, and Stefan Stremersch. "Predicting the consequences of marketing policy changes: A new data enrichment method with competitive reactions." Journal of Marketing Research 54, no. 5 (2017): 720-736.

Keelin, Thomas W., and Bradford W. Powley. "Quantile-parameterized distributions." *Decision Analysis* 8, no. 3 (2011): 206-219

Keelin, Thomas W. "The metalog distributions." *Decision Analysis* 13, no. 4 (2016): 243-277.

Keelin, Thomas W., and Ronald A. Howard. "The Metalog Distributions: Virtually Unlimited Shape Flexibility with Bayesian Updating in Closed Form." (2021).

Kim, Sungjin, Clarence Lee, and Sachin Gupta. "Bayesian synthetic control methods." *Journal of Marketing Research* 57.5 (2020): 831-852.

Manchanda, Puneet, Peter E. Rossi, and Pradeep K. Chintagunta. "Response modeling with nonrandom marketing-mix variables." Journal of Marketing Research 41, no. 4 (2004): 467-478.

McCarthy, Daniel Minh, and Elliot Shin Oblander. "Scalable data fusion with selection correction: An application to customer base analysis." Marketing Science 40, no. 3 (2021): 459-480.

Petrin, Amil. "Quantifying the benefits of new products: The case of the minivan." *Journal of political Economy* 110, no. 4 (2002): 705-729.

Rossi, Peter E., Greg M. Allenby, Robert McCulloch. *Bayesian Statistics and Marketing.* John Wiley & Sons, Ltd, 2005.

Thiruchelvam, Loshini, Sarat Chandra Dass, Vijanth Sagayan Asirvadam, Hanita Daud, and Balvinder Singh Gill. "Determine neighboring region spatial effect on dengue cases using ensemble ARIMA models." *Scientific Reports* 11, no. 1 (2021): 1-9.

van Zoest, Vera, Frank B. Osei, Gerard Hoek, and Alfred Stein. "Spatio-temporal regression kriging for modelling urban NO2 concentrations." *International journal of geographical information science* 34, no. 5 (2020): 851-865.

Wang, Craig and Reinhard Furrer. "Combining Heterogeneous Spatial Datasets with Process-based Spatial Fusion Models: A Unifying Framework." https://arxiv.org/abs/1906.00364

Watson, Sam. “Geostatistical modelling with R and Stan.” https://aheblog.com/2016/12/07/geostatistical-modelling-with-r-and-stan/

Yang, Jeremy, Dean Eckles, Paramveer Dhillon, and Sinan Aral. "Targeting for long-term outcomes." *arXiv preprint arXiv:2010.15835* (2020).

References for Section 2

Aiken, E., Thakur, V., & Blumenstock, J. (2022, June). Phone Sharing and Cash Transfers in Togo: Quantitative Evidence from Mobile Phone Data. In Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (pp. 214-231).

Alvarenga, K., Burattini, B., & Perin, G. (2022). Is going digital the solution? Evidence from social protection (No. 50). Policy in Focus.

Aron, J., & Muellbauer, J. (2019). The economics of mobile money: Harnessing the transformative power of technology to benefit the global poor. Centre for the Study of African Economies.

Best-Rowden, L., & Jain, A. K. (2018). Longitudinal Study of Automatic Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(1), 148-162.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 1-15.

Dhawan, R. (2020). The Aadhaar Effect: Why the World's Largest Identity Project Matters. ORF Issue Brief No. 378, Observer Research Foundation.

Fenton, G., & Muyundo, T. J. (2022). What next for humanitarian logistics? In Humanitarian Logistics: Meeting the challenge of preparing for and responding to disasters and complex emergencies (p. 285). Kogan Page Publishers.

Garvie, C., Bedoya, M. A., & Frankle, J. (2016). The Perpetual Line-Up: Unregulated Police Face Recognition in America. Georgetown Law Center on Privacy & Technology.

Grother, P., Ngan, M., & Hanaoka, K. (2019). Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects. National Institute of Standards and Technology Interagency Report, 8280.

Hamm, J., Kohler, J. C., Gretton, C., & Blease, C. (2021). Facial recognition technology and the disabled: A review of health and safety risks. International Journal of Technology Assessment in Health Care, 37(1), e59.

Hernandez, K., & Roberts, T. (2020). Predictive Analytics in Humanitarian Action: a preliminary mapping and analysis. K4D emerging issues report, 33.

Holloway, K., Al Masri, R., & Yahia, A. A. (2022). Digital identity, biometrics and inclusion in humanitarian responses to refugee crises. ODI.

Holm-Nielsen, P.V., Furu, P., & Raju, E. (2023). The influence of cash assistance on the localisation agenda in Kenya's humanitarian sector. Jamba, 15(1), 1496. https://doi.org/10.4102/jamba.v15i1.1496

Jaspars, S., Murdoch, C., & Majid, N. (2022). Digital feast and famine. Retrieved from https://www.calpnetwork.org/wp-content/uploads/2022/08/Digital-Feast-and-Famine-07.06.2022.pdf

Jeong, D., & Trako, I. (2022). Cash and In-Kind Transfers in Humanitarian Settings: A

Review of Evidence and Knowledge Gaps. Working Paper. Washington, DC: World

Bank. https://doi.org/10.1596/1813-9450-10026

Klare, B. F., Burge, M. J., Klontz, J. C., Vorder Bruegge, R. W., & Jain, A. K. (2012). Face Recognition Performance: Role of Demographic Information. IEEE Transactions on Information Forensics and Security, 7(6), 1789-1801.

Lallemant, D., Soden, R., Rubinyi, S., Loos, S., Barns, K., & Bhattacharjee, G. (2017). Post-disaster damage assessments as catalysts for recovery: A look at assessments conducted in the wake of the 2015 Gorkha, Nepal, earthquake. Earthquake Spectra, 33(1_suppl), 435-451.

Madianou, M. (2019). The biometric assemblage: Surveillance, experimentation, profit, and the measuring of refugee bodies. Television & New Media, 20(6), 581-599. https://doi.org/10.1177/1527476419857682

Martin, A., Sharma, G., Peter de Souza, S., Taylor, L., van Eerd, B., McDonald, S. M., ... & Dijstelbloem, H. (2023). Digitisation and sovereignty in humanitarian space: Technologies, territories and tensions. Geopolitics, 28(3), 1362-1397. https://doi.org/10.1080/14650045.2022.2033834

Mozur, P. (2019). One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority. The New York Times.

Nassiry, D. (2018). The role of fintech in unlocking green finance: Policy insights for developing countries (No. 883). ADBI Working Paper.

Owino, B. (2020). Harmonising data systems for cash transfer programming in emergencies in Somalia. Journal of International Humanitarian Action, 5(1), 11.

Paddock, J. (2021). Shared decision-making in humanitarian cash transfers (Master's thesis, Fordham University).

Pinna, D. (2020). Digitalisation of humanitarian cash aid. [Master's thesis, Wagemingen University]. https://edepot.wur.nl/532643

Raji, I. D., & Buolamwini, J. (2019). Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. AAAI/ACM Conference on AI, Ethics, and Society.

Tesfaye, W. (2017). AI and Mobile Cash Transfers in Zimbabwe: Improving Efficiency and Reach. Journal of Humanitarian Aid and Relief.

Pol, C. N. (2018). Smart cities solutions for refugee camps: Communication systems review to improve the conditions of refugees [Master's thesis, KTH Royal Institute of Technology]. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-238672.

Pisa, M., & Juden, M. (2017). Blockchain and Economic Development: Hype vs. Reality. Center for Global Development Policy Paper, 107.

Tirivayi, N., Matondi, P., Tomini, S. M., Mesfin, W., Tesfaye, S. C., & van den Berg Morelli, C. (2016). Humanitarian Assistance through Mobile Cash Transfers. United Nations University, 1-98.

Twigt, M. (2023). Doing refugee right(s) with technologies? Humanitarian crises and the multiplication of "exceptional" legal states. Refugee Survey Quarterly, hdad020. https://doi.org/10.1093/rsq/hdad020

UNICEF. (2023). East Asia and Pacific Region Humanitarian Situation Report (End-Year) 01 January to 31 December 2023. Retrieved from https://www.unicef.org/media/151721/file/East-Asia-and-Pacific-Region-Humanitarian-SitRep-(End-Year)-01-January-to-31-December-2023.pdf

Ussher, L., Ebert, L., Gómez, G. M., & Ruddick, W. O. (2021). Complementary currencies for humanitarian aid. Journal of Risk and Financial Management, 14(11), 557. https://doi.org/10.3390/jrfm14110557

Weitzberg, K., Cheesman, M., Martin, A., & Schoemaker, E. (2021). Between surveillance and recognition: Rethinking digital identity in aid. Big Data & Society, 8(1), 20539517211006744. https://doi.org/10.1177/20539517211006744

Zheng, et al. (2022). Discussion on AI-driven data quality improvement in humanitarian settings.