NEXUS™: ADVANCED MACRO-DYNAMIC MODELING AND OPTIMIZATION OF INTERNATIONAL MIGRATION

Utilizing Web-Scraping with NLP, Agent-Based Models + Network Analysis, and Labor-Market Analysis Inclusive of Behavioral Economics, Parameterized by Multi-Group CFA Data

ID: **S-BEHA-036(1651)T**

Tarun Batchu (Olentangy Liberty High School)

Non-Provisional Patent Application Status: Submitted

PROBLEM STATEMENT

HOOK: Every year, thousands of migrants risk their lives in search of safety, economic opportunity, or a better future. In 2024 alone, over **10,000 people died** attempting to reach Spain by sea.

- Beyond these tragedies, human smuggling networks thrive
- Because, migrant workers cannot obtain legal work permits, modern slavery continues to trap thousands in forced labor, as seen in the UK, with **over 19,000 cases of human trafficking were reported in 2024**.

These deaths and tragic accidents are the direct result of ineffective and restrictive migration policies: Lack of multilateral cooperation leads to fragmented policies, burdening frontline countries while others avoid responsibility.

- Lack of legal pathways →smuggling, misalignment with economic demands, lack of integration policies → anti-immigrant sentiment, failure to support circular migration.

Predictive Analysis: Data modeling can significantly improve migration policy, making it more effective, proactive, and humane.

- Gravity Models are of the most widely used tools but they fail to predict policy driven migration patterns and are **not inclusive**
- Other models are **low-scope and are not socially aware**

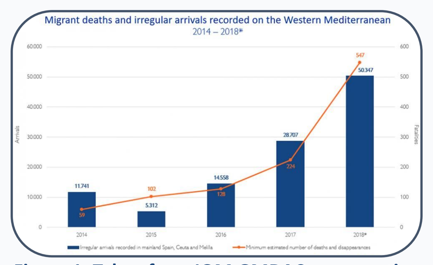


Figure 1. Taken from IOM GMDAC representing increasing migrant deaths recorded on the Western Mediterranean.

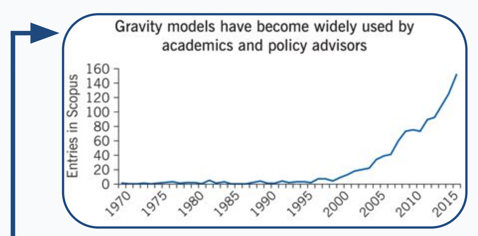


Figure 2. Modified from IZA. Schematic representation of growing usage of gravity models

$$\log\left(\overline{M}_{ij}\right) \sim p_0 + p_1 \cdot \log\left(\overline{P}_i\right) + p_2 \cdot \log\left(\overline{P}_j\right) + p_3 \cdot \log\left(\overline{I}_i\right) + p_4 \cdot \log\left(\overline{I}_j\right) - p_5 \cdot \log(D_{ij})$$

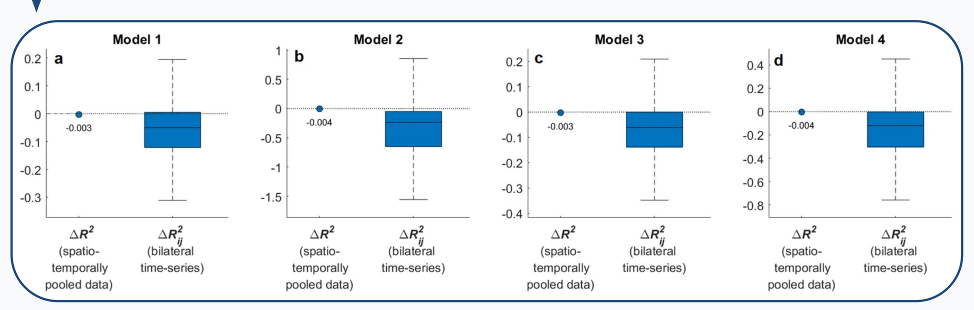


Figure 3. Modified from Beyer et al., 2022. Using modelled and observed temporal changes in migration flows at the bilateral level to show that gravity models do not explain, and cannot predict, international migration dynamics

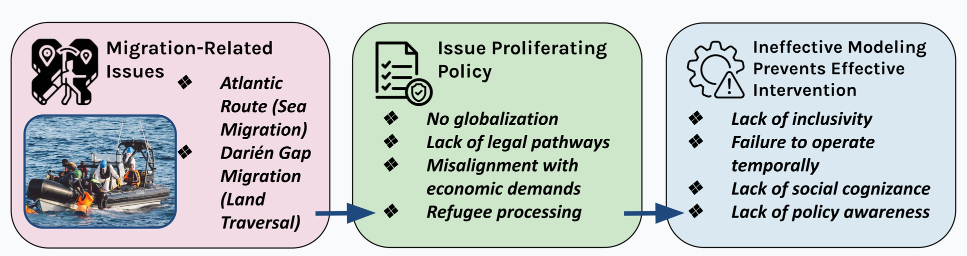
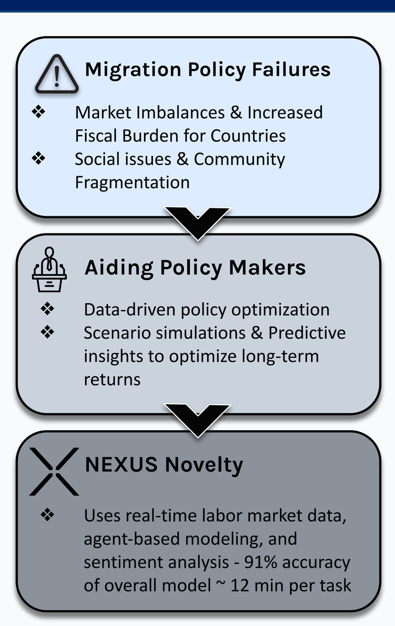


Figure 4. Collective representation of problem flow in international migration and policy refinement

VISUAL ABSTRACT



The inadequacies in migration policies have led to devastating impacts for those who leave their homes in the hope of upward mobility.

- These policies shaped by political short-termism lead to flaws within the international migratory process.
- This often materializes through deficits for both countries and migrants.
  - It is for these reasons that the International Organization for Migration finds that "in 2023, at least 8,565 people died on migration routes worldwide."

Because migration bolsters world economies and is considered a universal human right, policymakers must be well-informed in their decisions. This project aims to answer the question

- "Can a multi-modular machine learning model effectively optimize quotas?"
- On balance, our algorithm predicted migration with an R^2 value of 0.91 and simulated Monte Carlo decisions with a confidence interval of 99%. When evaluated using real data sets at 225,000 episodes, our aggregate model achieved an accuracy of **92% in identifying collective changes that could have been made to mitigate the conflict caused by the quota.**

RESEARCH OBJECTIVES & GOALS ☆

ENGINEERING PROBLEM: Poor regulation of international migration and corresponding policy proliferates thousands of deaths each year while leading to economic losses for all parties; predictive models can aid in the creation of more effective policy in a globally coordinated models but lack **accuray & inclusivity**

PROJECT GOAL: Promote upward mobility for all parties involved while mitigating structural oppression.

Engineering Design Criteria for Large Scale Data-Modeling Tool

Robustness: train models on World Bank, national statistic offices, OECD, UNHCR, IOM, Europol as well as social media platform data

Functionality: hardcoded to understand decision theories for ABM's and implement behavioral economics in QuantEcon

Inclusivity: use natural language processing for policy document analysis and web-scraping. Using confirmatory factor analysis and ARIMA.

Module specific research goals are expressed in the final report

CLAIM / RESEARCH QUESTION: Can the utilization of Agent-Based Models (ABMs) and Network Analysis alongside econometric and labor-market analysis with behavioral economics in a natural language processing web-scraping multi-modular machine learning model optimize immigration quotas or refine existing policy.

HYPOTHESIS: The utilization of the aforementioned tools implemented into a multi-modular machine learning model will on average, propose a policy modification.

ENGINEERING METHODOLOGY AND PROCEDURALS

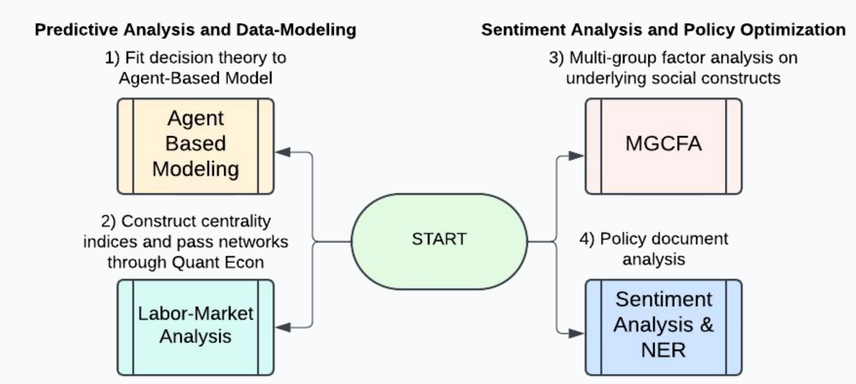


Figure 6. Condensed visual representation of multi-modular design of NEXUS. *Note: A full flow chart will be available on the flow chart.*

NEXUS leverages **Python, JavaScript, R, and Julia** to optimize migration policies through predictive modeling, labor market analysis, and agent-based simulations.

- **Python** powers **machine learning (scikit-learn, XGBoost), NLP (spaCy, transformers), and web scraping (Scrapy, Selenium)** to analyze migration trends, policy documents, and social sentiment. It also uses **Mesa** for agent-based modeling and **Factor Analyzer** to detect social biases affecting migration.
- **JavaScript** (React.js, D3.js) creates interactive dashboards for policymakers, while **Node.js + Puppeteer** enhances web data extraction.
- **R** handles **labor market evaluation** using **plm, tidyverse, and RStan**, integrating **QuantEcon** for economic modeling and policy forecasting.
- **Julia** enables **high-speed simulations (Agents.jl)** and **policy optimization (JuMP.jl)** to determine ideal migration quotas and workforce strategies.

Agent Based Modeling (ABM'S)

These **computational simulations** used to model complex systems by representing individual entities—called **agents**—that interact within an environment.

NEXUS employs **utility-based, network-driven, and behavioral agent-based models (ABMs)** to simulate migration flows, policy effects, and labor market dynamics.

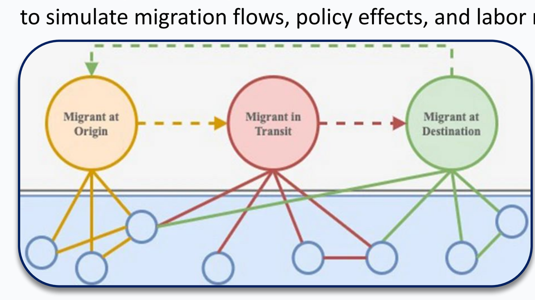


Figure 7. Adopted by McAlpine et al. 21. Framing low-wage labor migration as a complex dynamic system across multiple geographic spaces: Agents, networks and exogenous influencers

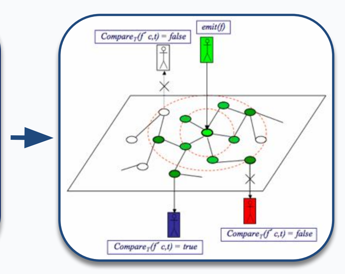


Figure 8. Sample schema exemplifying an environment mediated form of interaction

In **ABM**, decision theories define how agents **process information and make decisions**. The following are used in NEXUS.

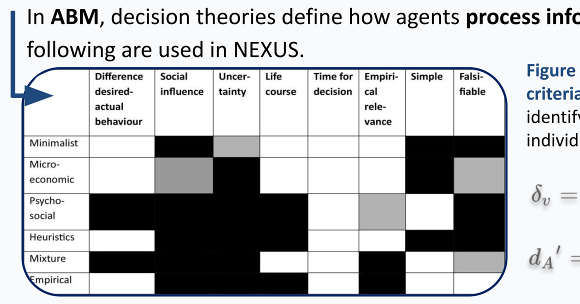


Figure 9. Fitting of decision theory to evaluation criteria from Klabunde 16. The evaluation aims to identify theories that best capture the complexities of individual migration decision-making.

$$\delta_v = \frac{|v_A - v_B|}{v_A + v_B}. \quad 1)$$

$$d_A{}' = d_A d_B + (1 - c_i)\, d_A t_B + c_u \delta_v t_A t_B, \quad 2)$$

Doubt Models

Network Analysis

Network analysis models international migration as a weighted-directed graph, where countries are nodes and migration flows are edges. This approach captures the interconnected nature of migration rather than treating countries as isolated entities.

- Network analysis is used to assign **burden responsibilities** to countries. Using **short-range (SRIC) and long-range (LRIC) interaction centrality indices**, the model will quantify direct and indirect migration influences, allowing for **data-driven policy optimizations**. Data: United Nations Population Division: 2022 Revision was used to construct the network. Modeled after **network theory developed by (Faud et al., 2016)** but with updated data.

Number of Countries by	Datasets			
	v2015		v2022	
	Inflows	Outflows	Inflows	Outflows
Citizenship	36	37	44	46
Residence	43	44	52	57
Place of Birth	1	–	1	–

Table 1 Distribution of countries by country of origin criterion. Characterization of data to train the model on building networks.

Burden Responsibility & Optimization

$$B_j = \sum_{i \in V} w_{ij} \cdot C_i \quad C_i = \frac{\sum_{k \in V} w_{ik}}{P_i}$$

B_j represents total burden responsibility for country j. C_i represents criticality. P_i is total population of country i. A country with a higher B_j value absorbs more migration pressure

SRIC & LRIC *Note: Equations will be broken down in the full report*

$$f(b, w_a) = \frac{p_{ba} + p'_{ba}}{|w_a|}, \quad c_{ij} = \begin{cases} \dfrac{a_{ij}}{\min_{\Omega(i) \subseteq N_i, j \in \Omega_p(i)} \sum_{l \in \Omega(i)} a_{il}}, & \text{if } j \in \Omega_p(i) \subseteq N_i, \\ 0, & j \notin \Omega_p(i) \subseteq N_i, \end{cases}$$

Policy Optimization Functions

To optimize migration policies, we define an objective function that minimizes migration burden **while ensuring equitable distribution**. Each country has a threshold. **Humanitarian Obg. Below**

$$\min \sum_{j \in V} \left(\frac{B_j}{P_j}\right)^2 \rightarrow \sum_{j \in V} B_j \geq \alpha \sum_{i \in V} \sum_{j \in V} w_{ij} \; ☆$$

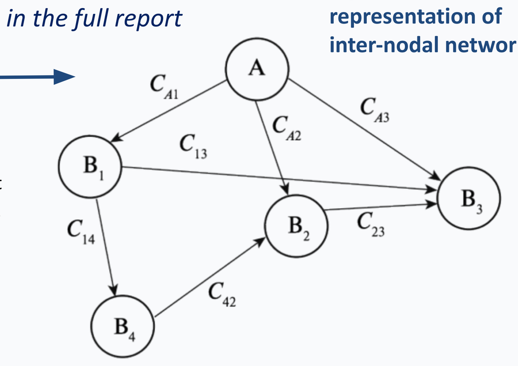


Figure 9. Schematic representation of inter-nodal network theory

Labor Market Analysis + Behavior Economics

Economic prosperity is a cornerstone of push-pull factors of migration: It is a major reason why immigrants are turned down at the border and why governments must make difficult decisions. Studying econometrics and labor-market analysis is crucial for optimizing and maximizing migration returns

Taraseyev '18:

- ***Positional game theory of individuals depending on economic factors***

1) $x^k_{ij}(t+1) = x^k_{ij}(t) + \dfrac{\alpha^k_j (M^k_i - x^k_i(t))(w^k_j(t) - w^k_i(t))}{W^k_j(t)}$

2) $w_j(t) = w_j(t_0) \cdot \dfrac{E_j(t_0) + V_j(t_0)}{E_j(t_0) + U_j(t_0) + x_j(t)}$

3) $w_i(t) = w_i(t_0) \cdot \dfrac{E_i(t_0) + V_i(t_0)}{E_i(t_0) + U_i(t_0) - x_i(t)}$

IZA '05:

- ***Neoclassical concept → demand and supply factors. Modeling Probability***

$$P = 1 - \Phi(z)$$

Yazar '23:

- ***Game theory considerations for smuggling***

$$\frac{\partial u_{SF}}{\partial \theta} = p_s v'_S((1-\theta) w_S + \theta w_D)(w_D - w_S) - (1 - p_s - p_k) v'_S((1-\theta) w_S + \theta \rho w_S)$$

$$(1-\rho) w_S - p_k v'_S((1-\theta) w_S) w_S = 0$$

Self Constructed:

- ***The model described herein incorporates the idea of interconnectedness and modeling with network analysis***

$$\max \frac{\delta U}{\delta \theta_{IA}} = p_{SI} \cdot v\left[U_{DC}(\lambda) + \theta_{BF} \bullet \psi_{MN}\right] - v \cdot \left[(1 - p_{SI}) \bullet (U_{DC} - U_{SC})\right] - \delta_{MSD} - \mu_{MPI} = 0$$

g→central network utility of intermigration

QuantEcon→ **open source libraries used for econometric analysis**

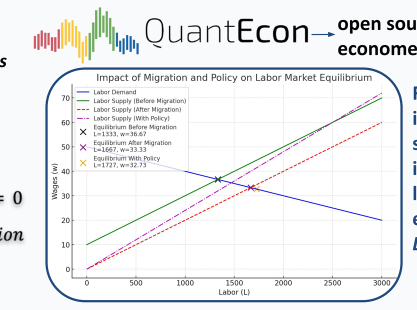


Figure 10. This figure illustrates an example simulation of the impact of migration on labor market equilibrium. *Market Dynamics*

SENTIMENT ANALYSIS

Multi-Group Confirmatory Factor Analysis (MGCFA):

NEXUS employs an (MGCFA) to identify latent social constructs shaping attitudes toward migrants. By analyzing survey data, media sentiment, and social network trends, NEXUS uncovers **key factors like economic anxiety, cultural distance, and social cohesion.**

Correlation Matrix[a]

		Economic Anxiety	Cultural Distance	Social Cohesion	ERPs	CIIs	CES
Correlation	Economic Anxiety	1.000	-.128	-.677	.126	-.802	-.887
	Cultural Distance	-.128	1.000	-.306	-.853	-.259	-.035
	Social Cohesion	-.677	-.306	1.000	.556	.911	.903
	ERPs	.126	-.853	.556	1.000	.365	.206
	CIIs	-.802	-.259	.911	.365	1.000	.924
	CES	-.887	-.035	.903	.206	.924	1.000
Sig. (1-tailed)	Economic Anxiety		.318	.002	.321	<.001	<.001
	Cultural Distance	.318		.125	.000	.166	.449
	Social Cohesion	.002	.125		.013	.000	.000
	ERPs	.321	.000	.013		.082	.223
	CIIs	.000	.166	.000	.082		.000
	CES	.000	.449	.000	.223	.000	

a. Determinant = 2.721E-5

Figure 11. MGCFA example Correlation Matrix conducted on a migration network (Philippines–Gulf States migration corridor) to find underlying factors behind social norms.

Sentiment Analysis NLP (Web-Scraping + Policy Document Analysis)

- Web scraping in **NEXUS** gathers real-time data from media. Tools like **Scrapy**, **Selenium**, and **Newspaper3k** extract content, while NLP techniques such as **Sentiment Analysis**, **NER**, and **Topic Modeling** analyze public sentiment and migration trends.
- To analyze policy documents and hardcode certain concepts on how policy is constructed, **NEXUS** uses **Large Language Models (LLMs)** like **BERT** due to their ability to understand and generate human-like text.

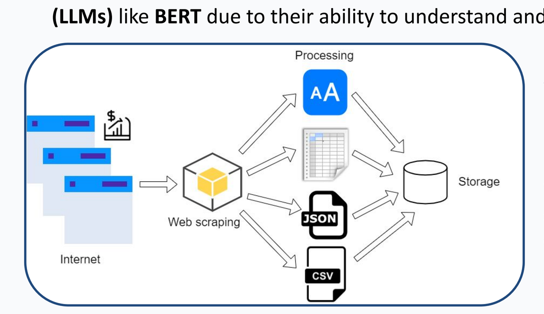


Figure 12. Visual for Natural Language Processing Web scraping developed by Ali '21

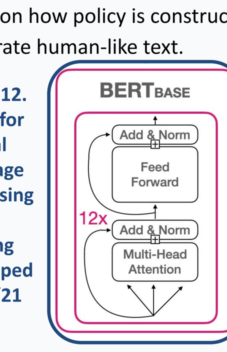


Figure 13. BERT model size and architecture for BERTbase. Used for understanding context for sentiment analysis.

We replace the standard **ReLU activation function** with **Swish** for smoother gradient flow to fine-tune BERT. We then perform **domain-specific pretraining** on migration-related texts. A **policy-aware loss function** is used.

$$\text{Consistency Loss} = \frac{1}{m} \sum_{j=1}^{m} (\text{Cosine Similarity}(\text{Output}_j, \text{Historical Policy}_j))$$

RESULTS & IN-SILICO VALIDATION

This project is developed with our personal machines (M1 macbook which is an ARM-based processor) and remote access to the high performance computing cluster in **Ohio Wesleyan University**. Data from IAS, IMA, OECD, World Bank, UN migratory services, UNHCR, Eurostat, Kaggle, and more are used. Simulations for NEXUS were conducted approximately 120,000 times over 1,150 epochs, incorporating countries at varying stages of development and networks of diverse sizes. These simulations totaled around 400 hours of computational time. The model utilizes 72 distinct input variables and generates 30 output metrics. **NEXUS is trained and validating using a minimization of harm criteria derived from economic harm reduction, social stability metrics, and political stability measures, while considering risk=benefit trade offs.**

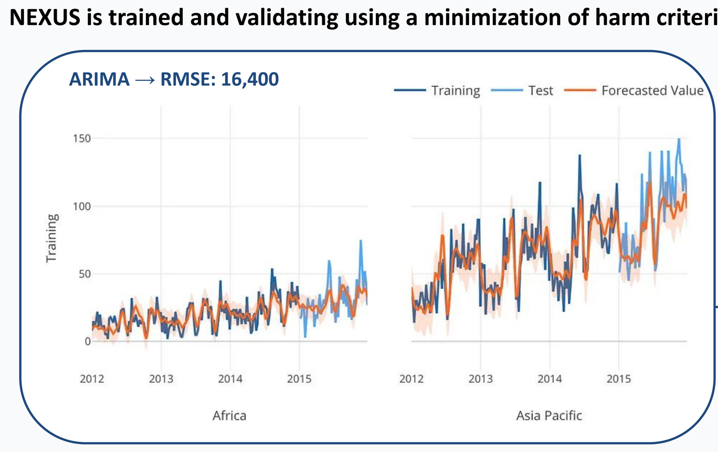


Figure 14. The chart compares actual and forecasted migration trends for Africa and Asia Pacific (2012–2015) using an ARIMA model. The model's RMSE is 16,400, indicating forecast accuracy and deviations from actual migration patterns.

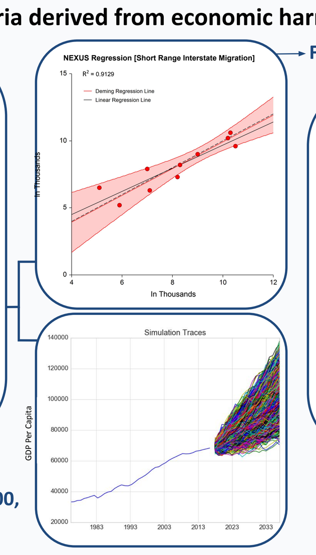
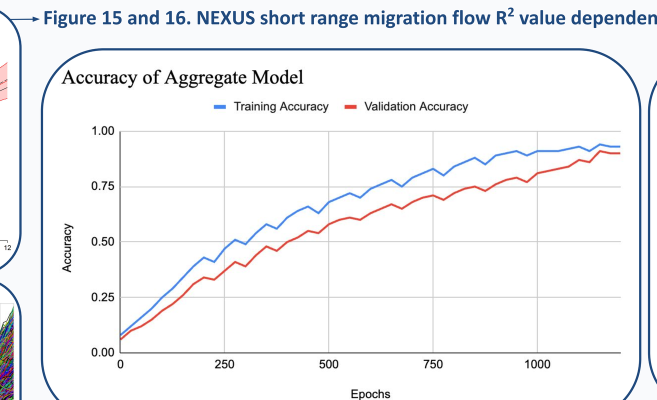


Figure 15 and 16. NEXUS short range migration flow R^2 value dependent on policy decisions.



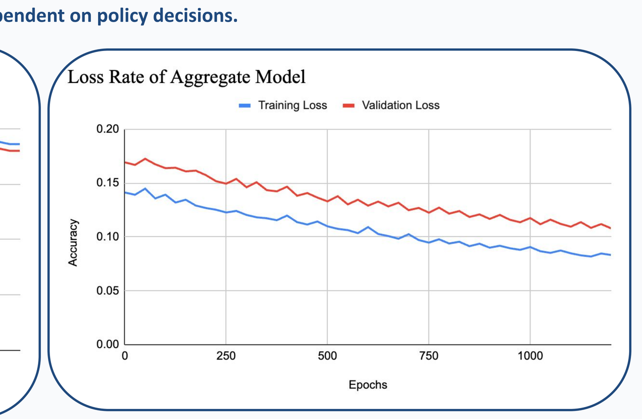


Figure 17 and 18. Accuracy and loss rate of aggregate model. This is relative to all the individual modules and the minimization of harm criteria using policy refinements.

DATA ANALYSIS

NEXUS leverages a multi-method analytical framework integrating **Agent-Based Modeling (ABM), econometric labor market evaluation, network analysis, and sentiment analysis** to optimize international migration policies.

- The **ARIMA model** (Figure 14) assesses example migration trends for Africa and the Asia Pacific, achieving a Root Mean Squared Error (RMSE) of **16,400**, indicating deviations between actual and forecasted migration flows. **Short-range migration analysis (Figures 15 & 16)** uses regression modeling to determine how policy changes influence migration rates, demonstrating a **strong R^2 value** in predicting migration shifts under varying conditions. **This is significantly better than current gravity models because they fail to model short-range bilateral migratory flows.**
- (Figure 17) show steady improvements over 1,150 epochs, with training accuracy surpassing 0.90 and validation accuracy converging. **Loss rate (Figure 18)** declines progressively, indicating **optimized parameter tuning** for policy decision-making.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad R^2_{adj} = 1 - \left(\frac{(1 - R^2)(n-1)}{n - k - 1}\right)$$

Econometrics

- *Labor-Market Absorption Rate*
- *Employment-to-Populati on Ratio*
- *Network Burden Assumes*

Social Cohesion

- *Inequality Index*
- *Social Cohesion Index*
- *Public Sentiment on Migration*

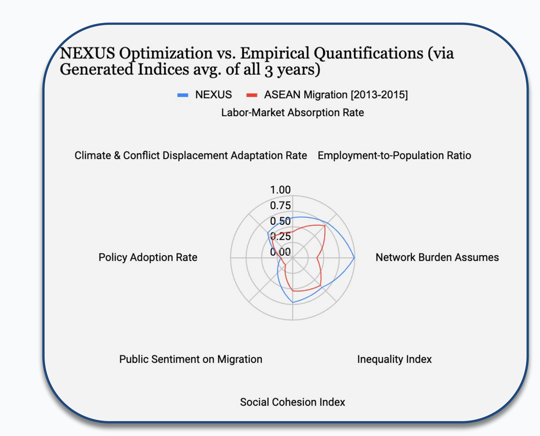
Policy Adaptability

- *Policy Adoption Rate*
- *Climate & Conflict Displacement Adaptation Rate*

A harm mitigation criteria is developed. After the model proposes policy recommendations, a sub-model generates indices based on the aforementioned criteria. This allows for evaluation of NEXUS refinements on broad spectra.

What Do Policy Recommendations Look Like?

For Economic Growth & Labor Markets. 1) Adjusts migration quotas based on labor shortages. 2) Prioritizes high-demand sectors with streamlined visas. 3) Encourages cyclical migration through incentives. **National Security & Border Stability.** 1) Predicts migration surges for proactive resource allocation. 2) Optimizes asylum processing using risk assessments. **Globalization & Economic Integration** 1) Develops regional labor mobility agreements. 2) Supports startup visas for entrepreneurs. **Social Equity & Migrant Mobility** 1) Matches migrants with optimal employment and housing. 2) Supports skill recognition for better job access. **Climate Migration & Sustainability** 1) Forecasts climate-driven migration for proactive policies. 2) Aligns migration with green job sectors.



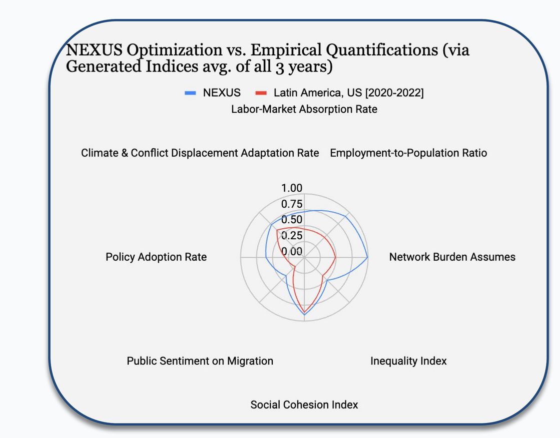


Figure 19 and 20. Harm mitigation testing of NEXUS general optimization. Radar chart showing visual of optimized criteria indices for two migratory networks (ASEAN [2013-2015] and Latin America, US [2020-2022]

CONCLUSIONS & FUTURE WORKS

The NEXUS model demonstrates a data-driven **socially aware** approach to optimizing migration policy by integrating predictive analytics, labor market evaluation, and sentiment analysis. By refining policies based on economic, social, and political stability metrics, the model offers targeted recommendations to minimize harm and promote upward mobility. Average wait time for the model per country given a specific network is approx. 734 seconds and is highly variable and dependent on country connections. **Currently, I am collaborating with the French Foreign Ministry to further develop NEXUS's applications in real-world policy settings.**

Model	Strengths	Limitations	NEXUS ☆
Macroeconomic Forecast Models (OECD, World Bank)	Provides historical economic insights and long-term projections.	Lacks real-time adaptability; does not consider social factors	Integrates real-time labor market data and sentiment analysis
Rule-Based Policy Systems (EU Blue Card, U.S. H-1B)	Structured migration pathways for skilled workers.	Rigid, leading to labor market mismatches. Lacks flexibility	NEXUS can adjust migration policies dynamically based on skill shortages
Humanitarian & Crisis Response Models (UNHCR, IOM)	Addresses urgent refugee and humanitarian crises.	Does not optimize for long-term economic integration.	Balances humanitarian response with long-term cohesion.
Traditional Agent-Based Models	Simulates migration under predefined conditions.	Often limited to static rules; lacks integration with real-time data.	Real-time feedback loops, optimizing policies on evolving conditions.

A well-designed UI is essential for NEXUS because policymakers need **clear, intuitive, and actionable insights** without requiring technical expertise. Migration policies involve complex economic, social, and political factors, so the interface must present **data-driven recommendations in an accessible format**.

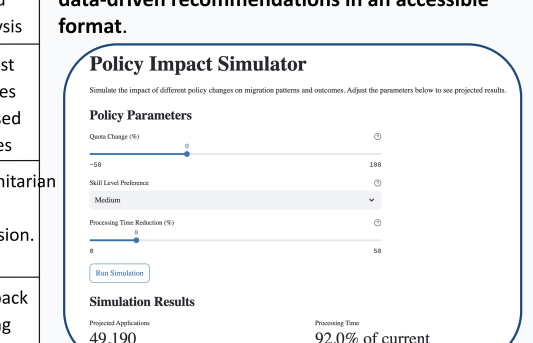


Figure 21. Example User Interface for Policy Makers

KEY REFERENCES

[1] Opportunities and Challenges of International Migration for Sending and Receiving Countries / T.C. Dışişleri Bakanlığı, https://www.mfa.gov.tr/opportunuties-and-challenges-of-international-migration-for-sending-and-receiving-countries.tr.mfa

[2] Putnick, D.L. and Bornstein, M.H. (2016) Measurement invariance conventions and reporting: The State of the art and Future Directions for Psychological Research, Developmental review : DR. U.S. National Library of Medicine. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5145197/ (Accessed: February 15, 2023).

[3] ArXiv.org e-Print Archive. https://arxiv.org/pdf/2104.02395.pdf.

[4] Alzubaidi, Laith, et al. "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions - Journal of Big Data." SpringerOpen, Springer International Publishing, 31 Mar. 2021, https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8.

[5] Council on Foreign Relations. (n.d.). The U.S. Immigration Debate. Council on Foreign Relations. Retrieved February 15, 2023, from https://www.cfr.org/backgrounder/us-immigration-debate-0

[6] Beyer, R. M., Schewe, J., & Lotze-Campen, H. (2022). Gravity models do not explain, and cannot predict, international migration dynamics. Humanities and Social Sciences Communications, 9(1). https://doi.org/10.1057/s41599-022-01067-x

I would like to extend a special thanks to Dr. Yazar, Dr. Hurst, Dr. Flory, Dr. Streib, Daniel Zakari, Aayush Varma, and Prasanjit Sahoo