Teodor Yankov

📞 +447470030723 | ✉ teodor.yankov@outlook.com | 🔗 My LinkedIn | My Portfolio

EDUCATION

University of Oxford — Oxford, UK
Master of Science, Social Data Science (High Merit) — *Oct. 2023 – Aug 2024*

- Courses in **Python**, analytical statistics, machine learning, network science, data analytics at scale, fairness and AI ethnics
- Thesis developing a novel NLP algorithm for computational narrative extraction (CNE) from unstructured text data using an ensemble of LLMs and signed Leiden community detection (outperformed existing CNE models on accuracy & recall tasks.

University College London (UCL) — London, UK
Bachelor of Science, Politics and International Relations with Data Science (First Class Honours) — *Sept. 2020 – May 2023*

- Courses in **R**, causal/statistical inference, measurement, quant. data analysis, migration, democracy, public policy, ethics.
- Dissertation exploring the predictive capacity of the **gravity model** in forecasting transnational refugee flows (strong correlations (up to 0.9) between actual and predicted flows, statistically significant at all conventional levels.)

Ongoing Certification Courses — Remote
Various Certified Providers — *Sept. 2020 – Present*

- `Sustainability` YOUNGO Nature Working Group: Biodiversity Academy, Oxford School of Climate Change, Climate Science for Everyone (Duke), Climate Data Analysis with Python (UK Centre for Hydrology and Ecology)
- `ML/AI` Databases and SQL for Data Science with Python (IBM), Introduction to SQL (Codecademy)

EXPERIENCE

Data Scientist `ML/AI` `Sustainability` `Research` `Product Dev.` — June 2022 – Present
World Data Lab — *Remote*

- Full-cycle contributor to data products **measuring** and **modelling** climate change, migration, spending and demography.
- **Curating** and **transforming** data into **digestible, decision-driving insights** for public (e.g. IOM, GIZ, Internet Society) and private (e.g. Mastercard, L'Oreal, Colgate) sector stakeholders on topics of sustainable development.
- Built **automated pipeline** with an ensemble of APIs to **gather** and **validate** datasets, simulating web search functionality in the OpenAI API to estimate product category market sizes.
- Spearheading **data insights mining** and **content creation** within **cross-functional teams** for high-impact global events on sustainable development and spending with participants such as Melinda Gates and Elton John.
- Led the **COP27 launch campaign** of the **World Emissions Clock**, presenting the cutting-edge climate monitoring tool to non-data science audiences and senior stakeholders.
- **Skills**: R, Python, data curation, analysis, modelling and visualisation, evaluation design, data product development.

Graduate Researcher `ML/AI` `Research` — July 2024 – Present
Oxford Internet Institute — *Oxford, UK*

- Lead author on ongoing research collaboration with Dr Adam Mahdi, Dr Scott Hale, Dr Luc Rocher and the University of Oxford's Centre on Migration, Policy and Society (COMPAS), building a **novel benchmark** for NLP algorithms using **synthetic ground-truth** via LLM-generated articles simulating discourse in the UK migration debate.
- Constructed efficient **web-scraper** to curate unstructured news article data beyond what available API tools could access.
- **Designing** the generative and evaluation architecture of the benchmark, involving **responsible fine-tuning** and **prompt engineering** of GPT-4, Mistral and Llama models through **ICL** to **mitigate risks** of unsafe and insensitive LLM output.
- Conducting **extensive model evaluation**, probing for **discriminability, human mimicry and hallucinations**.
- **Skills**: Python, data curation and refinement, NLP, LLMs (fine-tuning, prompt-engineering, ICL), evaluation design.

Research Assistant `Research` `Policy` — May. 2021 – Sept 2023
UCL Department of Political Science — *London, UK*

- Assisted Dr Samer Anabtawi, researching LGBTQ+ activism in the MENA region (working with **sensitive qual. data**).
- Worked on various projects engaging with **refugee experiences in unsafe settings** (e.g. Moria and Cox's Bazar refugee camps), **handling sensitive data** pertaining to testimonies, prioritising **safe research practices**.
- **Skills**: research ethics, data collection and processing, policy research, literature review.

Summer Intern - Data Research `Research` `Policy` — Jul 2021 – Sept 2021
Institute for Market Economics — *Sofia, Bulgaria*

- Synthesised reform plans submitted to the European Commission, led **refinement** of government & tax database for the assembly of regional profiles, used statistical software to filter and analyse **large-N EU funding data**.
- **Skills**: database curation and refinement, cross-functional teamwork and communication, handling private data.

Fellowships, Awards and Leadership

Global Shaper `Sustainability` `Leadership`   Sept 2022 – Present
World Economic Forum   *London, UK*
- Selected to join the WEF Global Shapers community in recognition of demonstrated leadership and social impact.
- Led projects on responsible AI use in job-seeking and the digital upskilling of refugees and asylum seekers in London.

Millennium Fellow `Sustainability` `Leadership`   Apr 2021 – Nov 2021
United Nations Academic Impact   *London, UK*
- Chosen for competitive (6% acceptance rate) fellowship from 25,500+ students.
- Founded the Connected Scholars Program - a 7-week speaker series empowering youth action towards the UN Sustainable Development Goals (100+ participants, £100+ raised).

Secretary & Welfare Officer `Policy` `Leadership`   Apr 2022 – May 2023
UCL United Nations Association (UCLUNA)   *London, UK*
- Organised highly successful event series, bringing together UCL societies and students with internationally renowned speakers across the UN and world governments for DEI workshops in the international governance space.
- Organising large scale Model United Nations conferences tackling the issue of minority empowerment, working with UNHCR & recognised by the UN Secretary General's Youth Envoy and the Secretary-General of the IOM.

Awards
- Dean's List for Outstanding Academic Achievement
- Winner and Team Leader - *Microsoft AI for Good Hackathon*
- *Wings of Hope Achievement Award* - awarded for promoting education for disadvantaged children in India & Malawi. Presented at the House of Lords on my work and the UN SDGs.
- *Owlypia TIC Challenge in Technology* (Gold)
- *UKMT Maths Competition* (Gold and Silver)
- *LIMUN Diplomacy Award* - awarded for my work in leading debate in various simulations of UN organs (e.g. UNSC, UNEP, UNHRC) on topics such as environmental refugees, female empowerment and terrorism.

Published Work and Contributions

`Published` Caballero Reina, J., Crespo Cuaresma, J., Fenz, K., Zellman, J., Yankov, T., Taha, A. **Gravity Models for Global Migration Flows: A Predictive Evaluation**. Popul Res Policy Rev 43, 29

`Published` Yankov, T. (2022) **The Digital Gap Strikes Again: Exploring the Effect of Internet Poverty on Refugee Flight Patterns Through Regression Analysis**. International Public Policy Review. 16(1), pp. 61-78.

`Published` Kharas, H., Fengler, W., Sheoraj, R., Vashold, L., Yankov, T., **Tracking Emissions by Country and Sector**. Brookings Institution.

`Published` Yankov, T. (2021) **Once Again Forgotten: Exploring the Disproportionate Impacts of COVID-19 on Rohingya Refugees with Future Implications for International Cooperation**. International Public Policy Review. 15(1), pp. 91-101.

`Forthcoming` Yankov, T., McNeill, R., Hale, S., Rocher, L., Mahdi, A. (2024) **MigNar : An AI-driven Synthetic Ground Truth Benchmark for Narrative Extraction Algorithms.** [Forthcoming]

`Forthcoming` Yankov, T., Mahdi, A. (2024) **NarrAI: Leveraging NLP, LLMs and Network Community Detection for Enhanced Narrative Extraction in the UK Migration Debate** [Forthcoming]

`Forthcoming` Crespo Cuaresma, Caballero Reina, J, J., Zellman, J, Fenz, K, Yankov, T. (2022) **Income projections for migration research: a framework based on human capital dynamics** [Forthcoming]

`Accepted at Conference` Yeung, J., Schuetze, S., Yankov, T. (2024) **Parallel Network Change: An Analysis of Migration-Trade-Terrorism Co-Evolution with Temporal Graph Distances and Latent Space Modelling**. Accepted at the *Networks and Time II* conference hosted by the *Network Science Institute* at *Northeastern University*.

Skills

Human Languages: English, Bulgarian, French, Spanish, German
Computer Languages: Python (proficient), R (proficient), SQL (basic), HTML (basic), CSS (basic)
Technical Skills: Full-cycle data manipulation (dataset curation, data analysis/visualisation, pipeline engineering), Machine Learning (multivar. linear/logit, regularised, tree-based and count regressions, neural networks), analytical statistics (hypothesis testing, causal inference), NLP (NER, topic modelling), SLMs/LLMs (GPT, Llama, Mistral, Claude, BERT, SpacY, fine-tuning, in-context learning, prompt engineering), network methods (community detection, temporal network analysis, multiplex and signed networks, network formation models, graph machine learning)
Tools / Libraries: TensorFlow, PyTorch, Numpy, Pandas, NetworkX, Scikit-Learn, Matplotlib, Transformers, tidyverse, data.table, GitHub, VS Code, R Studio, HuggingFace, Docker, GCP, LaTeX, Figma, Notion, WordPress, Google Suite
Soft Skills: presenting technical findings to non-technical stakeholders, cross-functional teamwork, time-management, leadership
Hobbies: painting, reading, bird-watching, swimming, skiing