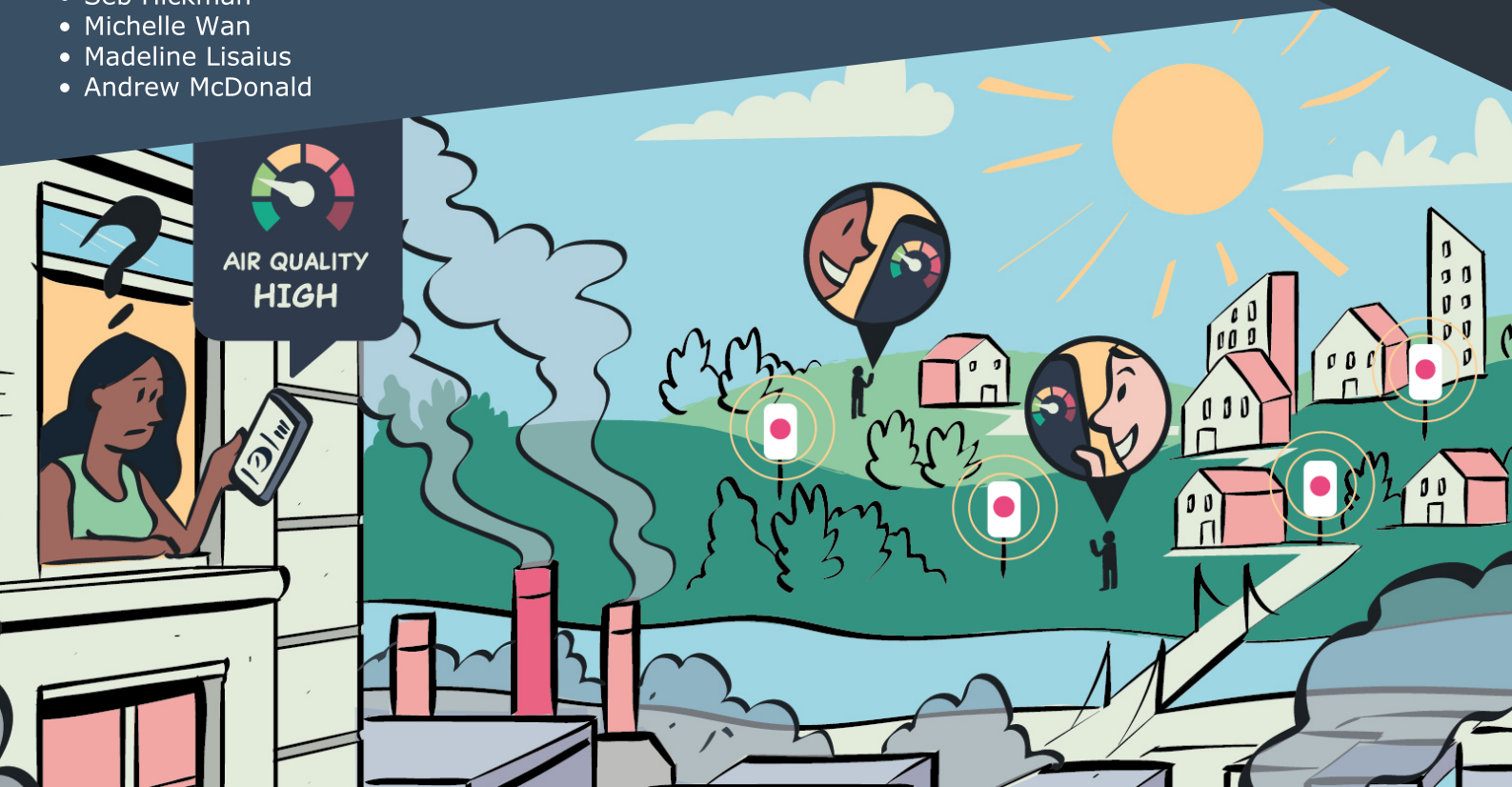
Air quality forecasting with an AI-driven system

Authors:
- Seb Hickman
- Michelle Wan
- Madeline Lisaius
- Andrew McDonald

The Alan Turing Institute



Overview

The Air Quality Group is a rapidly growing startup based in the UK that uses a black-box machine learning algorithm to forecast short-term air quality metrics with high fidelity. Goverments and organizations globally are interested in implementing this system.

Project Description

Air quality forecasting is an important tool for measuring and mitigating the negative health impacts from pollution exposure. Since pollutant concentrations can vary substantially across time and space, using tools to forecast these changes can help to guide immediate decision making as well as longer term planning.
The Air Quality Group (AQG) is a UK-based startup whose mission is to inform communities and decision makers about air quality. It has now developed a high fidelity, short term forecasting tool for levels of particulate matter (PM2.5), ozone (O3), nitrogen oxides (NOx), and sulphur dioxide (SO2). This forecasting tool uses a black-box machine learning model which has been trained using large quantities of UK air quality and meteorological data.
AQG's work has attracted the attention of governments worldwide, particularly in countries badly affected by air pollution. These parties are keen to use the modelling system as quickly as possible, to inform responsive and preventative decision-making. An initial internal review by AQG has identified the following areas of concern:

- Scientific accuracy of the model
- Downstream social & policy implications resulting from action taken based on model predictions
- Profitability of an open-source business model

Technology Description

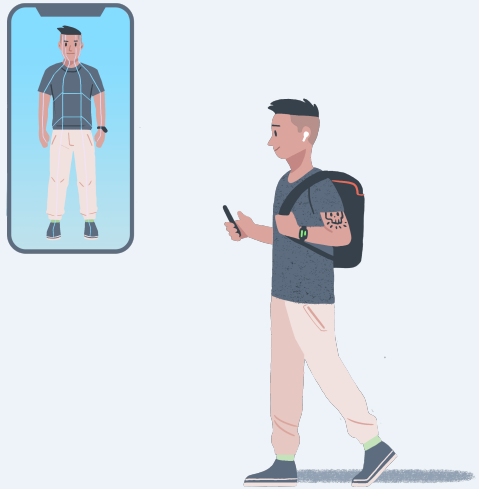
AQG has created a deep learning model which is trained on UK air quality data from UK-AIR and meteorological data from the Met Office. The model was built using a black box neural network architecture, and has not been designed with explainability in mind. The algorithm outputs hourly forecasts of multiple air pollutants at high spatial resolution, using substantial compute resources to achieve these predictions.

Key Issues

- **Scientific accuracy and generalisability of the model across varying climates, contexts, and geography**
- **Accountability for policy and social implications resulting from decisions taken based on imperfect model forecasts**
- **Business case supporting the profitability of the model – an open-source model allows others to interrogate the algorithm, and drive improvements, but limits profitability**
- **Explainability of the black-box model**
- **Sustainability and scalability of compute-intesive forecasts from large models in carbon-intensive data centres**
- **Supporting counterfactual forecasts to compare policy interventions**

Deliberative prompts

1. **Who is responsible for inaccurate predictions? To what extent is AQG accountable for the downstream impacts of inaccurate air quality forecasts?**
2. **AQG has trained a model exclusively on data from the UK. What implications might this have when other governments use it?**
3. **What issues may arise if this short-term forecasting system is used to make longer term forecasts?**
4. **How can AQG balance scientific transparency with competitive advantage? Should the AQG be required to open-source the model to develop its explainability and interpretability? How might this erode profitability or expose security risks to bad actors?**
5. **How can AQG manage model uncertainties, and their communication? Should model development favour over- or under-prediction?**



Datasheet

Category	Details
Available Data	• UK air quality data for PM2.5, O3, NOx, SO2. • UK meteorological (weather) data including temperature, relative humidity, precipitation, wind speeds and direction
Analysis Techniques	• Deep neural networks for timeseries prediction

Groups, Organisations and Affected Individuals

1. **Air Quality Group startup**
2. **Governments**
3. **Affected populations (residential and workplace exposure)**
4. **Major polluters**