Fairness metrics in AI and national income: A Policymakers dilemma?

Shikhar Mishra

September 4, 2023

1 Background

Themes of **algorithmic bias** from artificial intelligence algorithms are currently a hot topic of debate among academics and policymaking circles. For example, the COMPAS software system is now actively assisting the American judiciary in assessing the suitability for parole based on chances of recidivism.

The **bias-variance tradeoff** is a well documented statistical phenomenon. It can be understood as a 'no-free-lunch' principle; reducing bias from (say) a predictive model comes at the cost of increased variance in the model, and vice versa. Precision and accuracy are often used interchangeably with bias and variance.

Technological improvements (in the wide-reaching economic sense) hold a central position in **endogenous growth theory**, popularised through the works of Romer. The technological capacity of a state is a key force multiplier in the economic production function. This can be the (marginal) difference between a developing nation 'catching-up' or not.

2 The dilemma

Lets say the state utilises its fiat and regulates by evaluating AI algorithms on fairness metrics in various sectors of the economy. For example, lets say the normative policy is a ban on using a given demographic characteristic, say gender, by banks in determining credit-card applications. Assuming complete compliance means that gender is not longer an input feature into, nor a quasi-input (via inferring from other features) into the final decision model. In this limiting case the variable "Gender" will contribute zero variance into the explainability of the decision to accept or reject by the bank.

In the case where the banned variable did not contribute significantly to final decision, there would be no impact. But in the cases where did contribute, and since the decision space of the model is no longer allowed partitioning on the basis of gender, ceterus paribus means there is a loss in model accuracy with respect to binary choice to reject or accept. However from a fairness perspective this is a more 'fair' outcome. The exact quantification will vary from one fairness metric to another, but the directionality is the same.

To address the reduced variance in their classification model, banks can choose to build a more accurate algorithm (without using the banned variable), possibly by adding other 'approved' features. The banks which find the best features will thrive, while those that make less accurate credit approval decisions will go bust. The business cycle will continue.

The state creates incentives (and disincentives) for firms by regulating different sectors of the economy with preference or fiat for 'fairer' algorithms. Some sectors may be able to handle the impact of algorithmic regulation better than others. The sectors which compete globally are most sensitive to domestic regulations, this includes firms in the high-tech economy.

More research can be conducted to quantify how a fairer algorithm regime impacts a nations endogenous technology factor, but intuitively two countries with different regulatory AI regimes will affect their technological factor of growth proportionately.

It is important to make distinction between the perspectives of state and firms. Firms are motivated by (mostly) monetary profits, but the state has a responsibility to be fair and ethical. The cost of positive 'fairness' effects is traded off with negative monetary effects of regulation (eg. by capital or innovation flight). All regulations create these deadweight losses and economies have correspondingly lower economic output.

A regulatory regime could have a **two-pronged solution**, one for official state or societal algorithms and another for the private sector. The state and societal algorithms have a need to be fair, but can be less accurate. While the firm created algorithms should be allowed the R&D and space to be more accurate, sometimes at the cost of fairness.

The dilemma for the policymaker of a nation is that it is not as simple to just approve more fair or ethical algorithms, they come at the cost of global competitiveness and also a medium to long-run tradeoff from living standards by inhibiting the tech factor. This is the pareto frontier which the policymakers could optimise on, and be transparent with the public on.

3 Resolution

'High' AI regulation regimes will attempt to coerce the 'low' AI regulation regimes into common compliance standards. The 'low' countries may choose to accept in return of equivalent benefits (monetary, military, or sentimental) or reject in favour of continuing developing their competitive advantage, or be promiscuous.

The other (not often discussed publicly) direction can also work to achieve equilibrium. If there is sufficient evidence to conclude that a 'low' AI regulation regime is not or will not attempt to increase its standards, the high regime can one-sidedly choose to reduce their regulation to match the lower standards. But this again comes with political costs. Society may wish to be relentless in a noble pursuit, without fully internalising (or understanding) all the costs of the new state of affairs.

The author believes that the state (and its advisors) should endeavour to be omniscient and act in the best judgement of the nation, factoring in all evidence. Decision making should be dynamic enough, and without lag, to allow for swift reversals in state policies if new evidence arises.