This came about by going back to the definitions of what I am working on. I am going to define my main task as follows: How can we detect contrarian framing about the energy transition in Australia?

> **Framing**: The selection of certain elements of a perceived reality to make more salient in a communicating piece (Entman 1993)
> **Contrarianism**: An ideological stance unwavering in the face of evidence against it (Brisman 2012)
> **The Energy Transition**: The energy market readjusting to take into account the negative externalities of the historical methods of energy production (John Carlin 1995)

From these definitions, it is reasonable to say that people with contrarian views on the energy transition are opposed to this market readjustment. This is most often due to not believing the externalities are real or impactful, or the economic implications of this readjustment are too dire (*Note the parallels to the three main counter climate claims: it's not us, it's not bad, it's not worth solving (Brisman 2012)*). Their contrarian frames will go on to highlight these concerns while ignoring the positives of renewable energy.

Following this line of thinking, I want to reframe my original task as: What does an article say about the economic or external impacts of a piece of energy infrastructure in comparison to the *'truth'* as defined by the broader discourse. This operationalises framing in such a way that large scale analysis is possible, while also preserving the nuance that will be required for later interventions.

Theoretical Method

A piece of communicating text will have any number of claims. Each claim is a judgment on a specific type of energy infrastructure and should be simple enough to be summarised into a single adjective (*for example: 'Nuclear Power' – 'Expensive', 'Wind Turbines' – 'Deadly'*). From these adjectives, two features are then extracted: dimension and valence. There are two dimensions, economic or external (*for example: 'Expensive' – Economic, 'Deadly' – External*) and valence scores the magnitude and direction of the adjective respective to this dimension between -1 and 1.

By combining all the claims from a large set of texts, a baseline distribution can be constructed which new texts can be compared against.

Baseline Distribution
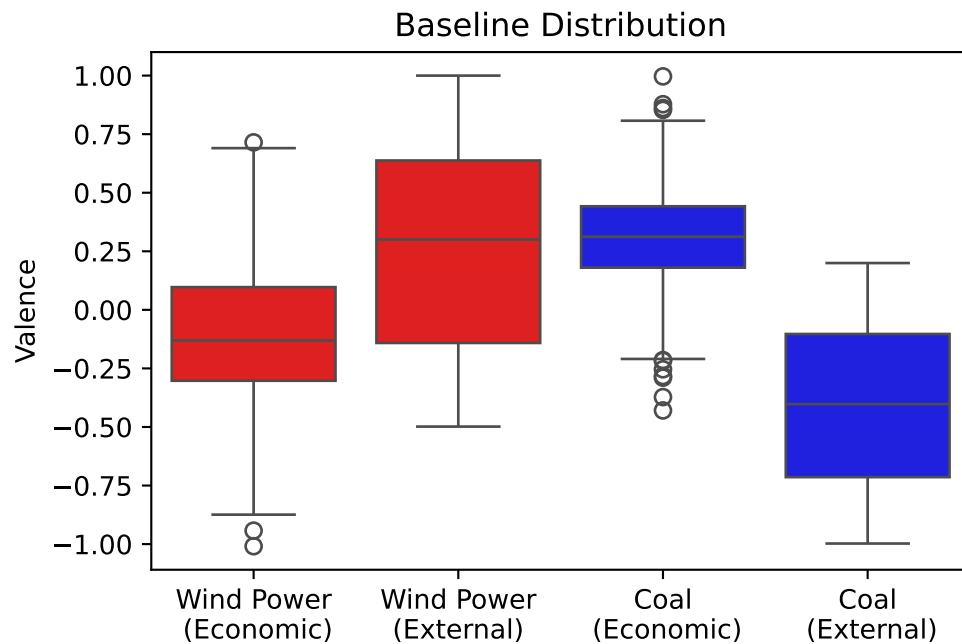
If it seems that the distribution of claims is probable, then the text wouldn't be flagged. This is shown in the chart below:
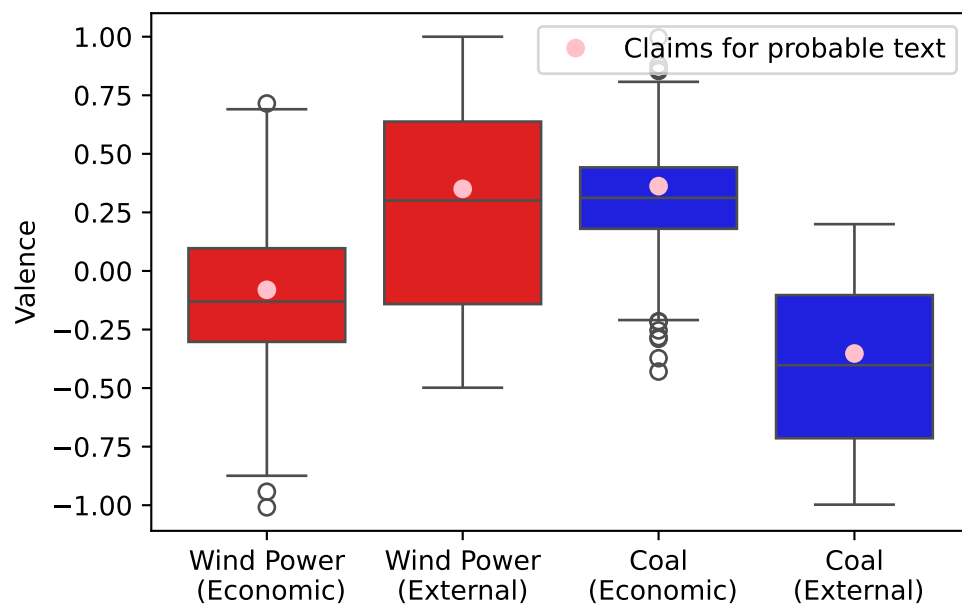
However, if the claims do not fit the distribution, they would be flagged as contrarian if they are biased. If the claims are unbiased, but the distribution says that they should be, they would be flagged as false balance.
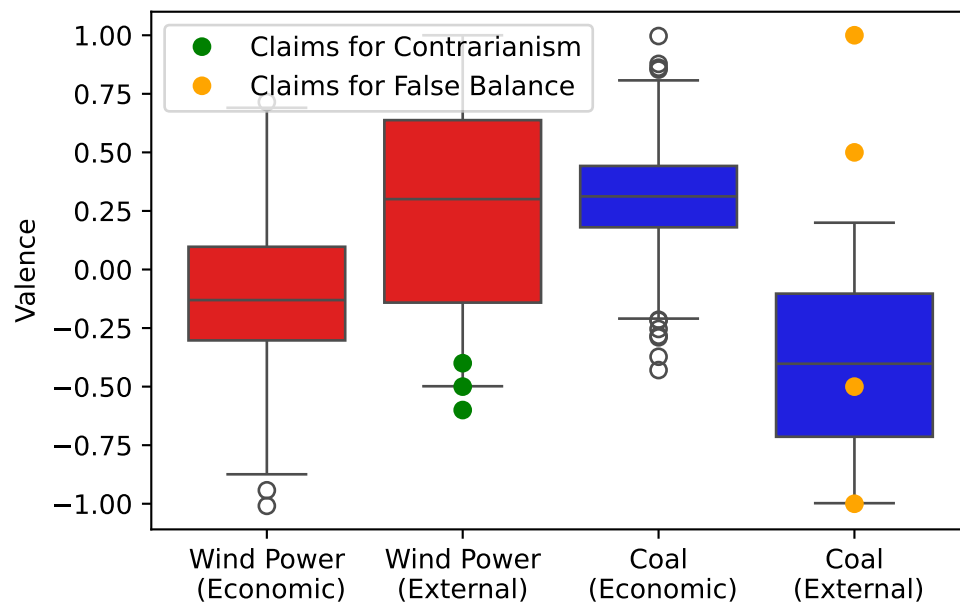
After the flagging, the evidence for an intervention can be constructed by gathering complementary claims from the baseline distribution that would pair with the claims in the text to make it better fit the distribution.
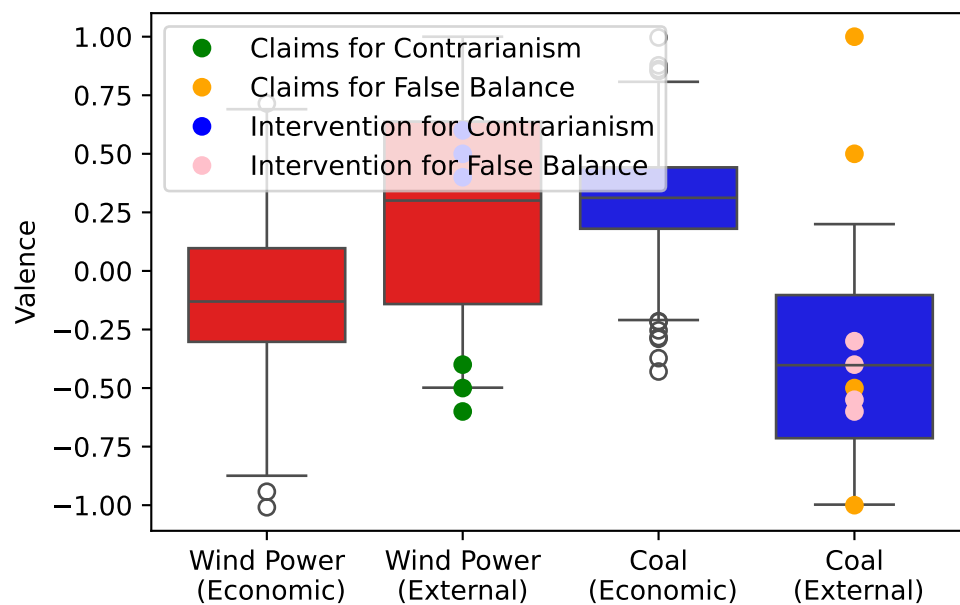
Implementation

From my prototyping, GPT-4 has been good at performing the entire task of claim extraction in one step; however, I am experimenting with separating a few of the tasks to reduce cognitive load. I should have a demo running next week. I will probably use the media bias ratings to help establish what goes into the

baseline distribution, but that's a decision I'm making down the line. I'm also planning on doing a bit of labeling to have an evaluation system for the claim extractor.

Brisman, Avi. 2012. "The Cultural Silence of Climate Change Contrarianism." In *Climate Change from a Criminological Perspective*, edited by Rob White, 41–70. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-3640-9_4.

Entman, Robert M. 1993. "Framing: Toward Clarification of a Fractured Paradigm." *Journal of Communication* 43 (4): 51–58.

John Carlin. 1995. "Environmental Externalities in Electric Power Markets: Acid Rain, Urban Ozone, and Climate Change." U.S. Department of Energy. https://www.eia.gov/electricity/pdfpages/enviroext.pdf.