Generative AI Diversity Guidelines

Six Principles for Responsible Journalistic use of Generative AI and Diversity and Inclusion

Generative AI (GAI) programs, such as ChatGPT and Bing, are increasingly entering UK newsrooms as a tool for British journalists. The use of AI in journalism raises specific challenges when it comes to the issue of Diversity, Equity and Inclusion, and there are still ongoing discussions whether Generative AI can be used ethically and effectively in newsrooms. These guidelines are not to endorse the use of Generative AI in newsrooms but are intended to raise issues that should be considered with specific reference to diversity and inclusion **if** it is used.

The algorithms of Generative AI tools rely on processing large quantities of existing source materials. It is commonly acknowledged that existing British journalism suffers from a diversity problem with an over-representation of white men. For example in 2020 Women in Journalism published research showing that in one week in July 2020 - at the height of the Black Lives Matter protests across the world - UK's 11 biggest newspapers failed to feature a single byline by black journalist on their front pages. Taking non-white journalists as a whole, of the 174 bylines examined only four were credited to journalists of colour.

The same report also found that in the same week just one in four front-page bylines across the 11 papers went to women.

Importantly the week the study surveyed the biggest news stories were about Covid-19, Black Lives Matter, the replacement of the toppled statue of the slave trader Edward Colston in Bristol and the appeal over the British citizenship of the Muslim mother, Shamima Begum.

This means that assuming the algorithms of Generative AI programs draw on the stories written by journalists in mainstream newspapers to generate its information, if a journalist were to ask it any questions about the issues in the news that week they will overwhelmingly be receiving information from a white male perspective.

The end result is that Generative AI programs, if used inappropriately, will only serve to reinforce and amplify the current and historical diversity imbalances in the journalism industry effectively building bias on top of bias.

The lack of diversity and inclusion in the source material that Generative AI uses is of course not only limited to journalism but also applies to numerous other fields as well including the sciences and academia.

While we urge all Generative AI programmers and software designers to address these concerns, as well as urge media organisations (and other sectors of society) to improve their diversity and inclusion in order to increase the diversity of the source material, there are steps that all journalists can undertake right now to work in a more ethical and responsible manner when it comes to diversity and journalism.

We have proposed six basic media diversity principles that all journalists and media organisations should abide by. As Generative AI changes, and its use in newsrooms adapts, these principles should also change and be dynamic over time. We also do not see these six principles as definitive. Instead we see this as an urgent intervention to address the current lack of public discourse around this critical issue.

We actively encourage these six principles to be interrogated by practitioners and academics and for them to be built upon.

Six Basic Principles
1. **Be aware of built-in bias**
   Journalists and media organisations need to recognise the potential for bias inherent in the use of current Generative AI models when it comes to diversity. To be explicitly aware of an issue is always a critical step in addressing a problem, just as we are expected to be aware of the bias inherent in all our sources, whether because of vested interests or the limitations of personal experience. Once we are aware of built-in bias we can build on the same strategies that we use with human sources, e.g. careful questioning, background research, second-sourcing etc.
2. **Be transparent where appropriate**
   Journalists and media organisations should be transparent in their use of Generative AI when, and where, it is appropriate. What level of use of Generative AI in the production of a

piece is appropriate before declaring its use will depend on how it is used, change with time, and depend on the issues covered. This should be an ongoing discussion with the journalism industry, creating and promoting industry standards. At this point we would, at the very least, suggest that directly using text created by Generative AI should be clearly labeled. We would also encourage media organisations to publish their policies and guidelines around the use of Generative AI.

3. **Build diversity into your prompts**

   Ask for diverse experts and perspectives. Journalists should explicitly seek, through their prompts, for Generative AI to draw on source material written and/or owned by different demographics.

   Where this is not possible journalists should use prompts to obtain lists of experts and recognised commentators on specific issues from different backgrounds. Going to the original work of these experts and commentators directly can complement any material created by Generative AI and address possible biases.

4. **Recognise the importance of source material and referencing**

   Journalists should respect and acknowledge the work of the creators of content that Generative AI draws on to produce its results. Historically the lack of acknowledgement of original work has disproportionately fallen on people from under-represented and marginalized backgrounds. To achieve this, we would encourage journalists to use Generative AI programs that explicitly list the source material used in the creation of its text.

5. **Report mistakes and biases**

   All journalists have a responsibility to contribute to creating a better media sector and improve tools used by journalists. When biases are spotted and issues arise when using Generative AI programs journalists should report these to the programmer and software developers (this is often possible within the Generative AI tool through your own responses and/or the 'thumbs up/down' buttons). Similarly best practice should also be fed back in order for the programmers of Generative AI to build better models.

6. **GAI-generated text should be viewed with journalistic scepticism**

   Do not rely on Generative AI created text as an authoritative source of information GAI is well known for 'hallucinating' facts and other information in its responses, creating fictional individuals and sources. No information provided by GAI should be treated as fact, but be viewed as "informed plausibilities" — it is best used to provide suggestions that are then followed up for further exploration. We would also encourage journalists to use Generative AI programs that explicitly list the source material used in the creation of its text.

Conclusion

We recognise that there is the potential for the use of Generative AI in journalism to increase exponentially over time.

We believe that if news organisations and individual journalists use Generative AI they should view it as a tool rather than a replacement for journalists. We also believe that it is vital that if and when it is used it is used in a responsible way that can address related issues of media diversity or at the very least ameliorate some of the worst problems.

However, we also recognise that many of these problems are created by a lack of diversity on the source material in the first place - due to the under-representation of certain demographics in various different sectors from academia to the media, as well as how AI programmers choose and weight the source material that Generative AI algorithms use. Therefore, while these are six principles of how individual journalists should use Generative AI it is still incumbent on wider society to increase the diversity of their respective sectors and for Generative AI programmers to examine how they can also address diversity issues.

According to a survey by the World Association of News Publishers currently half of all newsrooms use Generative AI tools, yet only a fifth have guidelines in place, it is unclear if any of these guidelines explicitly address diversity and inclusion. This must be rectified as soon as possible.

Supplemental Note

Possible examples of Generative AI bias

1. On the 10 June 2023 when prompted; "Who are the twenty most important actors of the 20th Century?"
    ChatGPT did not name a single actor of colour
2. On the 13 June 2023 when prompted: "What are the important events in the life of Winston Churchill?"
    Bing failed to mention his controversial views on race, his controversial role in the Bengal famine, and his controversial views towards the Jews or Islam.

3. On the 10 June 2023 when prompted: "What are important facts about the American founding fathers?"
   Chat GPT failed to mention that any of them owned slaves.

We are not dictating, or even suggesting, that journalists should include these facts when covering these three issues. However it seems to clearly point to a certain perspective that traditionally would be thought does not represent the concerns and priorities of disproportionately historically marginalised groups.