

MPhil Politics, Comparative Government

YouGov Experimental Design

February 21, 2025

Contents

Experimental Pre-Analysis Plan	2
Research Question	2
Theoretical and Empirical Motivations	2
Hypotheses to be Tested	3
Confirmatory Hypothesis	3
Exploratory Hypotheses	4
Experimental Protocol	4
Treatment Conditions	4
Assignment	5
Measuring outcomes	5
Testing Hypotheses	5
Power Analysis	5
Assumptions	5
Calculations	5
Design Threats	5
Bibliography	6

Experimental Pre-Analysis Plan

Research Question

Does exposure to AI-generated news increase affective polarisation?

Theoretical and Empirical Motivations

Advancements in machine learning techniques, particularly transformer models trained to efficiently handle sequential data inputs and outputs, have popularised the field of Artificial Intelligence (AI) (Vaswani *et al.*, 2017). Amongst AI's applications, generating hyper-realistic textual and visual content has become easily accessible, helping AI become an enabling informational tool. Yet, as unregulated AI technologies remain prone to hallucinations and misuse from bad actors, they are raising concern in social and political contexts (Duberry, 2022; Rawte *et al.*, 2023). AI can be used to generate manipulative political information and deceitful deepfakes which can be used to incite hate or spread misinformation. Questions are therefore being raised on whether AI-generated content influences voting behaviour and election outcomes such that it poses a threat to the trust and integrity of democratic political institutions (Stockwell, 2024).

Structural effects of globalisation and economic liberalism, coupled with individual political failings and electoral shocks have created an increasingly unequal and divided world. Consequent disillusionment and disconnected identities have encouraged voter volatility and rising populist narratives, notably in the United Kingdom (UK) (Norris and Inglehart, 2019; Fieldhouse *et al.*, 2019: 28-32). This environment — coupled with social media — has aided the dangerous spread of fake news which has been shown to favour populists, affect voting behaviour, and strengthen identities and affective polarisation within echo chambers (Cantarella, Fraccaroli and Volpe, 2023; Pfister *et al.*, 2023; Hobolt, Lawall and Tilley, 2023). Despite minimal literature on AI in political science, early research suggests AI-generated messages can also be persuasive and compelling (Bai *et al.*, 2023; Goldstein *et al.*, 2024). But, when aware of political content being AI-generated, readers become sceptical of news veracity more generally, even if the content is true (Altay and Gilardi, 2024). With voters showing scepticism towards AI-generated content, Cashell (2024) argues creators are turning to deepfakes to perpetuate existing views and stereotypes rather than attempting to persuade new views. As AI-generated content can be compelling and may be used to polarise in similar ways to fake news, the volatile political landscape also provides fertile ground for widespread dissemination of deceitful AI-generated information.

My research question builds upon the rise of fake news, and fills a distinct gap in the new AI literature. In particular, my research question seeks to provide validity to Cashell (2024)'s arguments of AI being used to amplify views amongst

asks whether exposure to AI-generated news increases affective polarisation? This question seeks to understand the whether AI-generated news can influence attitudes towards opposite partisans

Given fears that AI-generated news can contribute to the spread of misinformation, as well as manipulate and spread hate through often indistinguishable deepfakes, AI-generated news is being associated with untrustworthy content (Altay and Gilardi, 2024). Veracity of the news article is an important factor to control and consider

I aim to investigate whether exposure to AI-generated news increases affective polarisation. This research is important as it will provide insights into how AI-generated news can influence voting behaviour and election outcomes, and whether it poses a threat to the trust and integrity of democratic political institutions.

As trust in news sources is key to how voters perceive and act on information for decision-making, the implication of AI-generated news diminishing trust in news sources is that it may also diminish trust in the political system and institutions, thus increasing polarised views on articles opposed to their core beliefs (Hobolt, Lawall and Tilley, 2023).

My explanatory hypotheses will test whether the fears that AI-generated news is associated with fake news are accurate, alongside testing the expectation that the veracity of an AI-generated news articles will not affect outcomes of trust

The research focuses on the UK to expand the literature beyond the United States. The dependent variables are conceptually grounded in voting behaviour and valence theory, with consideration given to their operationalisation and measurement validity so results can be reliably used for further research (Adcock and Collier, 2001; Goertz, 2006; Green and Jennings, 2012; Fisher, 2017). However, further literature review is required to better identify whether attitudes and affective polarisation may be susceptible to persuasion from a simple experimental exposure. Running a pilot study may be an effective way to provide credence to my initial hypotheses before refining the research design.

If AI is shown to influence affective polarisation, it could validate populists using the technology to shape political discourse and threaten institutions, risking democratic backsliding (Haggard and Kaufman, 2021). The implications of this research topic would inform how we regulate, highlight, or restrict AI-generated news — whether inaccurate or not.¹

Hypotheses to be Tested

Confirmatory Hypothesis

H_1 : Exposure to AI-generated news articles will increase affective polarisation.

¹However, aggregate-level effects of AI on the 2024 UK election were minimal (Simon, McBride and Altay, 2024).

Exploratory Hypotheses

H_2 : Exposure to political content labelled as AI-generated will decrease trust in the veracity of AI-generated news articles will not affect polarisation. H_3 : AI-generated news articles are associated with fake news.

Experimental Protocol

Treatment Conditions

Table 1: Treatments and Control for AI-generated News Exposure.

	News Article Labelling	
	No Labels	Labelled as AI
Control	Human-generated Article	N/A
Treatment	AI-generated Article	AI-generated Article

Note: Treatment variations to test for interaction effects of veracity, ideological stance, context, and source will be used.

1. Definition of the treatment conditions
2. How units are assigned to treatment conditions
3. How you will measure the primary outcomes.
4. How your primary hypotheses will be tested

Assignment

Measuring outcomes

Testing Hypotheses

Power Analysis

1. Defining and trying to justify the assumptions that guide your analysis and the range of parameters that you will consider.
2. Conduct a simulation-based power calculation using R.
3. Plot power curves to show how changing the values of key parameters impact the power of your study.

Assumptions

Calculations

Design Threats

1. Discuss whether there are potential issues of non-compliance, attrition, or spillovers in your design.
2. Discuss which measures will be taken to verify and minimise potential threats presented by non-compliance, attrition, or spillovers.

Bibliography

Adcock, R. and Collier, D. (2001) ‘Measurement Validity: A Shared Standard for Qualitative and Quantitative Research’, *The American Political Science Review*, 95(3), pp. 529–546.

Altay, S. and Gilardi, F. (2024) ‘People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation’, *PNAS Nexus*, 3(10), pp. 403–414.

Bai, H., Voelkel, J.G., Eichstaedt, Johannes C. and Willer, R. (2023) ‘Artificial Intelligence Can Persuade Humans on Political Issues’. OSF [preprint]. Available at: <https://doi.org/10.31219/osf.io/stakv>.

Cantarella, M., Fraccaroli, N. and Volpe, R. (2023) ‘Does fake news affect voting behaviour?’, *Research Policy*, 52(1).

Cashell, N. (2024) ‘AI-generated images: How citizens depicted politicians and society’, *UK Election Analysis*.

Duberry, J. (2022) ‘AI and information dissemination: Challenging citizens access to relevant and reliable information’, in *Artificial Intelligence and Democracy*. Cheltenham: Edward Elgar Publishing.

Fieldhouse, E., Green, J., Evans, G., Mellon, J., Prosser, C., Schmitt, H. and van der Eijk, C. (2019) ‘The Rise of the Volatile Voter’, in E. Fieldhouse, J. Green, G. Evans, J. Mellon, C. Prosser, H. Schmitt, and C. van der Eijk (eds) *Electoral Shocks: The Volatile Voter in a Turbulent World*. Oxford University Press, pp. 50–73.

Fisher, J. (2017) ‘Persuasion and mobilization efforts by parties and candidates’, in *The Routledge Handbook of Elections, Voting Behavior and Public Opinion*. Routledge.

Goertz, G. (2006) *Social science concepts: A user’s guide*. Princeton: Princeton University Press.

Goldstein, J.A., Chao, J., Grossman, S., Stamos, A. and Tomz, M. (2024) ‘How persuasive is AI-generated propaganda?’, *PNAS Nexus*, 3(2).

Green, J. and Jennings, W. (2012) ‘The dynamics of issue competence and vote for parties in and out of power: An analysis of valence in Britain, 1979–1997’, *European Journal of Political Research*, 51(4), pp. 469–503.

Haggard, S. and Kaufman, R. (2021) ‘Backsliding: Democratic Regress in the Contemporary World’, *Elements in Political Economy* [Preprint]. Available at: <https://doi.org/10.1017/9781108957809>.

Hobolt, S.B., Lawall, K. and Tilley, J. (2023) ‘The Polarizing Effect of Partisan Echo Chambers’, *American Political Science Review*, 118(3), pp. 1464–1479.

Norris, P. and Inglehart, R. (2019) *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*. Cambridge: Cambridge University Press.

Pfister, R., Schwarz, K.A., Holzmann, P., Reis, M., Yogeeswaran, K. and Kunde, W. (2023) ‘Headlines win elections: Mere exposure to fictitious news media alters voting behavior’, *PLOS ONE*, 18(8).

Rawte, V., Chakraborty, S., Pathak, A., Sarkar, A., Tonmoy, T.I., Chadha, A., Sheth, A.P. and Das, A. (2023) ‘The Troubling Emergence of Hallucination in Large Language Models - An Extensive Definition, Quantification, and Prescriptive Remediations’. arXiv. Available at: <https://arxiv.org/abs/2310.04988> (Accessed: 2 January 2025).

Simon, F.M., McBride, K. and Altay, S. (2024) ‘AI’s impact on elections is being overblown’, *MIT Technology Review*.

Stockwell, S. (2024) ‘AI-Enabled Influence Operations: Threat Analysis of the 2024 UK and European Elections’, *Centre for Emerging Technology and Security*. <https://cetas.turing.ac.uk/publications/ai-enabled-influence-operations-threat-analysis-2024-uk-and-european-elections>.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017) ‘Attention Is All You Need’, in *31st Conference on Neural Information Processing Systems*. Long Beach, CA, USA: arXiv.