University of Oxford

MPhil in Politics: Research Design in Comparative Political Science

Edward Anders

2025-02-25

Research Question

Does exposure to AI-generated news increase affective polarisation?

Theoretical and Empirical Motivations

Streucture of Theoretical Argument

- **The Rise of AI**: a powerful, but potentially dangerous tool
- **Voter Volatility**: a fertile ground for manipulation and polarisation
- **Gaps in Literature**: limited focus on attitudes and affective polarisation
- **Theoretical Argument**: how AI may influence affective polarisation
  - Associations between AI-generated news and fake news
  - Trust in news sources and political institutions
  - Veracity and partisanship linkages

Theoretical and Empirical Motivations I

The Rise of AI

- **The Transformer Model**: Advancements in machine learning have made AI more efficient at processing sequential data, enabling hyper-realistic content generation.
  - *(Vaswani et al., 2017)*
- **Accuracy Fears**: AI's accessibility as an informational tool raises concerns due to its susceptibility to hallucinations and misuse, including the spread of manipulative political content and deepfakes.
  - *(Duberry, 2022; Rawte et al., 2023)*
- **AI in Politics**: There is growing debate on whether AI-generated misinformation impacts voting behaviour and election integrity, potentially threatening democratic institutions.
  - *(Stockwell, 2024)*

Theoretical and Empirical Motivations II

Voter Volatility

- **Globalisation and economic liberalism**, alongside political failures and electoral shocks, have exacerbated inequality and division, fostering voter disillusionment and rising populism, particularly in the UK.
  - *(Norris and Inglehart, 2019; Fieldhouse et al., 2019: 28-32)*
- **Social media** has facilitated the spread of fake news, which disproportionately benefits populists, influences voting behaviour, and reinforces identity-based polarisation.
  - *(Cantarella, Fraccaroli and Volpe, 2023; Pfister et al., 2023)*
- **Echo chambers intensify affective polarisation**, strengthening group identities and deepening political divides.
  - *(Hobolt, Lawall and Tilley, 2023)*

Theoretical and Empirical Motivations III

Existing Literature: Effects of AI

- **Persusion and propaganda**: AI-generated messaging can be persuasive and compelling across many policy areas, although effect sizes are small.
  - *(Bai et al., 2023; Goldstein et al., 2024)*
- **Labelling Effects**: When aware of political headlines being AI-generated, readers can become sceptical of news veracity more generally, even if the content is true.
  - *(Altay and Gilardi, 2024)*
- **Expected Use of AI**: political use of AI is expected to focus on deepfakes to perpetuate existing views and stereotypes rather than attempting to persuade new views.
  - *(Cashell, 2024)*
- **AI's use in elections**: aggregate-level effects of AI on the 2024 UK election were minimal.
  - *(Simon, McBride and Altay, 2024)*

Theoretical and Empirical Motivations IV

Research Gap and Contribution

- **Limited in scope**: most research on the effects of AI are focussed on the US.
- **Policy Implications**: my research aims to provide additional insight on how we should regulate, highlight, or restrict AI-generated news. Labelling AI-generated news as AI-generated may have unintended consequences.
- **Understanding Effects**: providing a better understanding of whether fears of AI having negative effects on politics are justified.

Theoretical and Empirical Motivations V

Theoretical Argument

- **AI as fake news**: AI-generated content may polarise opinion in a similar way to fake news. Readers feel a greater sense of loyalty to their in-group and more distrusting of the out-groups as the fake news reinforces their existing beliefs.
  - *Azzimonti and Fernandes (2023)*
- **Fake news and trust**: when readers are aware that news is fake, they are less likely to trust news, politicians, and democracy.
- **Veracity and Partisanship**: readers associate veracity with the partisan narratives they want to believe, therefore helping polarise views.
- **Awareness of AI**: if readers are aware that the news is AI-generated, they may be less likely to trust the veracity of the news, the news source, the political institutions, and, importantly opposing partisans.
  - *(Altay and Gilardi, 2024)*

Hypotheses

Confirmatory Hypothesis

H_1: Exposure to AI-generated news articles will increase affective polarisation, but effect sizes will be small.

Exploratory Hypotheses

H_2: Exposure to political content labelled as AI-generated will decrease trust in the veracity of AI-generated news articles.

H_3: The veracity of AI-generated news articles will not affect polarisation.

H_4: AI-generated news is associated with narratives of opposing partisans.

Research Design I

To test these hypotheses, I will use a simple between-subjects experimental design. This randomised survey experiment will be conducted through the YouGov UniOM scheme with a representative sample of ~2,000 UK residents.

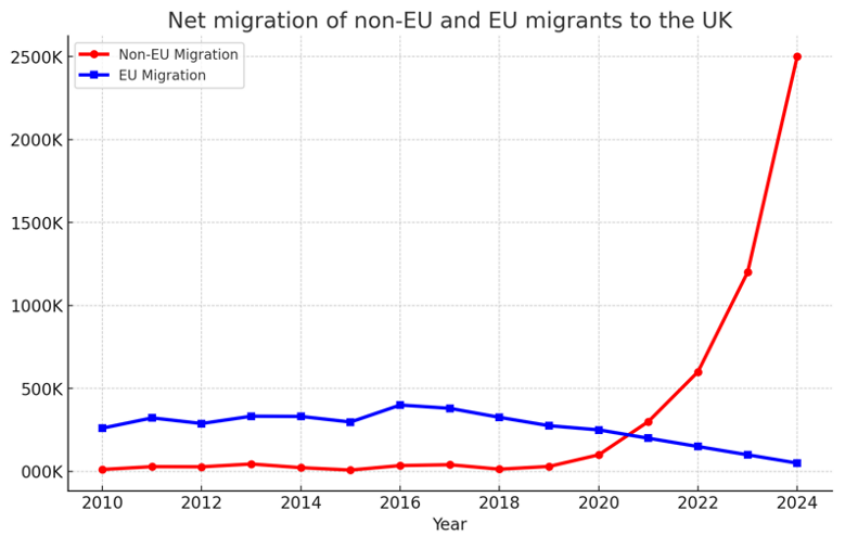
Table 1: Treatments and Control for AI-generated News Exposure.

	News Article Labelling	
	No Labels	**Labelled as AI**
Control	Human-generated Article	Human-Generated Article
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 in subsequent studies. Labelling dis-entangles the effects of exposure to the news content and the role of an article being AI-generated.

Treatment Example

BORDER CRISIS: NON-EU MIGRATION SKYROCKETS WHILE EU NUMBERS PLUNGE!



UK Overrun as Non-EU Migration Surges to Unprecedented Levels! Britain is facing an immigration crisis like never before, with non-EU migration spiralling out of control, dwarfing traditional European migration numbers, shocking new figures reveal. Official statistics show that while EU migration has steadily declined—even enjoying a brief peak around 2016—non-EU arrivals have exploded, soaring to eye-watering levels...

Research Design III

Treatment Assignment

- Two treatment conditions and one control condition.
  - Treatment: AI-generated news article (labelled or unlabelled as AI)
  - Control: Human-generated news article
- Participants will be randomly assigned to one of the three conditions in order for **Average Treatment Effects (ATE)** to be estimated due to randomisation.

Measuring Outcomes

- Primary dependent outcome variable to measure is affective polarisation.
- Post-treatment survey questions of both the control and treatment groups.
  - This is due to the temporal proximity of pre- and post-exposure survey questions so to avoid pre-test sensitisation effects.

Research Design IV

Measuring Outcomes

No.	Question	Options
1	**Affective Polarisation**	
1.1	How would you rate each of the following groups or people below? Keir Starmer, Nigel Farage; The Labour Party, Reform UK	Numeric thermometer scale: 0-100 (100 degrees = favourable and warm. Between 0 and 50 degrees = unfavourable and cold)
1.2	How comfortable are you having a political discussion with a $OTHERPARTY?	Extremely comfortable, Somewhat comfortable, Not too comfortable, Not at all comfortable
1.3	How much of the time do you think you can trust the $OTHERPARTY to do what is right for the country?	Almost never, Once in a while, About half the time, Most of the time, Almost Always

Note: Affective polarisation measures should use feeling thermometer ratings of the other party, trait ratings of the other party, whether or not respondents trust the other party, and a set of social distance items on the willingness to interact with opposing partisans [@levendusky2021; @druckman2019].

Research Design V

Hypothesis Testing

- To test the causal mechanisms of how exposure to AI-generated news affects affective polarisation, additional mediator variables are measured.

No.	Question	Options
2	**Trust in Media and Democracy**	
2.1	How much do you trust the news media to provide accurate information?	Likert scale: 1 (No trust at all) – 7 (Complete trust)
2.2	How much trust do you have in how democracy functions in the UK?	Likert scale: 1 (No trust at all) – 7 (Complete trust)

Research Design VI

Hypothesis Testing

- To test the causal mechanisms of how exposure to AI-generated news affects affective polarisation, additional mediator variables are measured.

No.	Question	Options
3	**Article Credibility and Influence**	
3.1	Do you think the claim in the headline you saw was true or false?	True / False
3.2	How trustworthy did you find the news article you read?	Likert Scale: 1 (Not trustworthy at all) – 7 (Extremely trustworthy)
3.3	Has the article changed your opinion on the topic?	Yes, positively; Yes, negatively; No change; Don't know

Threats and Limitations

Complex Treatment Interactions

- Dis-entangling the effects of the exposure to an AI-generated article.
- Are the outcome measures the result of AI, content, labelling, context, veracity, or source?
- Use of multiple treatment variations to test for interaction effects.

Attrition

- Possible dropouts from the study may lead to biased results.
- The survey questions may be too difficult or cumbersome to answer.
- Make the questions short and clear to answer.

Null Effects

- The effects of AI-generated news may be minimal or non-existent.
- I need to situate the experiment in as extreme scenario as possible.
- How long do effects actually last? Are there compounding effects resulting from repeated exposure?

Bibliography i

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.

Azzimonti, M. and Fernandes, M. (2023) 'Social media networks, fake news, and polarization', *European Journal of Political Economy*, 76, p. 102256. Available at: https://doi.org/10.1016/j.ejpoleco.2022.102256.

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) 'Al-generated images: How citizens depicted politicians and society', *UK Election Analysis*.

Bibliography ii

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.

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

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.

Bibliography iii

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*.

Spohr, D. (2017) 'Fake news and ideological polarization: Filter bubbles and selective exposure on social media', *Business Information Review*, 34(3), pp. 150–160. Available at: https://doi.org/10.1177/0266382117722446.

Bibliography iv

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.