University of Oxford: MPhil in Politics

Research Design in Comparative Political Science

1090063

Research Question

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). To better understand these concerns, my MPhil research question asks:

> *What are the causal effects of AI-generated news articles on voter attitudes of party and leader competence, issue ownership, and the economy?*

This question builds upon the rise of fake news, and fills a distinct gap in the new AI literature. 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 encouraged 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 propaganda produced by AI can be compelling (Bai *et al.*, 2023; Goldstein *et al.*, 2024). But, when aware of political content being AI-generated, readers become sceptical of its validity even if the content is true (Altay and Gilardi, 2024). Given possible scepticism towards veracity, Cashell (2024) argues deepfakes are used to perpetuate existing stereotypes rather than attempting to persuade new views. As AI-generated content can be compelling and 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. Therefore, my research hypothesis is that AI-generated content can influence political attitudes. Although not a primary research focus, possible mechanisms include effects on trust in politicians, media, and democratic institutions.

To test this hypothesis, my research investigates whether exposure to AI-generated articles affects political attitudes, and trust more generally. 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). Consequently, if AI is shown to influence attitudes, it could validate populists using the technology to shape political discourse and threaten institutions, risking democratic backsliding (Haggard and Kaufman, 2021). Whilst further DPhil research would address the mechanisms through which AI influences attitudes, and aggregate-level effects on elections, the implications of this research topic would inform how we regulate, highlight, or restrict AI-generated news — whether inaccurate or not.[1] This essay proceeds to evaluate appropriate research designs and their limitations for answering this research question.

Research Design

My research aims to identify the direction and size of *effects of causes*. Causal mechanisms of *how* and *why* effects occur is not explicitly within scope, but initial indications may be found. I make the principal argument for a between-subjects laboratory experimental research design by justifying large-N methods, before explaining the value of laboratory experiments for providing strong internal validity, and finally evaluating the limitations of this experimental design.

Small- vs Large-N Research Designs

To identify valid causal effects, the independent variable(s) must have isolated, exogenous variation that is independent of observed and unobserved confounders to ensure conditional independence. To then estimate causality through counterfactual comparisons, the positivity assumption should hold giving a non-zero probability that the treatment is received (Holland, 1986; Przeworski, 2009).[2] Although King, Keohane and Verba (1994) argue causal effects can be estimated using a unified 'logic of inference' across both qualitative and quantitative methods, small-N research design is not appropriate for estimating casual effects of AI-generated news articles on political attitudes. Small-N methods focus on comparing a small number of cases to generalise about the case's population, but small samples with limited variation encourages Omitted Variable Bias and endogeneity (Brady and Collier, 2010: 197). Whilst thick case analysis helps explain specific cases, small-N studies often focus on within-case analysis meaning conclusions cannot be postulated to wider populations without questions of selection bias and replicability (Goertz and Mahoney, 2012: 89). Alternatively, large-N designs use large, cross-case comparison with experimental or natural randomisation to achieve exogeneity, conditional independence, and positivity (Goertz and Mahoney, 2012: 102). A large-N design is therefore most appropriate for my research due to being able to leverage exogenous variation, apply statistical techniques for robust estimates, and isolate moderator variables.

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

[2] Unit homogeneity, Stable Unit Treatment Value Assumption (SUTVA), and no measurement error are implicitly assumed.

Experiments for Causal Inference

For rigorous causal empiricism, internal validity — achieved through a focused research question and strong identification strategy — is prioritised over generalisation and theoretical development. (Sammi, 2016: 942). In a model of causal inference, units u (e.g., an electorate's voters) are associated with an outcome response variable $Y \rightarrow Y(u)$, which is affected by a treatment t (e.g., AI-generated news articles) or a non-treatment control c (Holland, 1986: 945). The causal effect of t on the unit U, relative to the control is:

$$Y_t(u) - Y_c(u) \tag{1}$$

However, the Fundamental Problem of Causal Inference arises as it 'is impossible to observe the value of $Y_t(u)$ and $Y_c(u)$ on the same unit' (Holland, 1986: 947). A credible theory-led laboratory experiment — instead of simply large-N observations which favour statistical power — is best-suited to solve this problem (Titiunik, 2015). Experiments test a causal proposition through random assignment of conditions given to treatment and control groups to allow for counterfactual comparisons, independent of confounding variables (Druckman *et al.*, 2011: 4). In particular, laboratory experiments ensure randomisation and controlled settings to help satisfy the core conditions of causal inference for internally valid results (Druckman, 2022). Random assignment of the treatment ensures the cause is isolated and exogenous of any observed or unobserved confounders. Consequently, this exogeneity of the treatment guarantees conditional independence such that the causal effects are unbiased estimates, unaffected by external factors (Holland, 1986: 948). Randomisation also ensures unit homogeneity. As unobserved heterogeneity in characteristics are randomly distributed, units are assumed comparable for valid counterfactual analysis (Druckman, 2022: 29). However, my research is also interested in understanding which voter characteristics affect susceptibility to effects of AI-generated news. Based on UK literature, valence-based political attitudes are often correlated with views on integrity, leadership, past performance, image, and party management (Green and Jennings, 2012). Therefore, these heterogeneous covariates will be explicitly controlled in my analysis, with additional tests for moderator effects.[3] Another benefit of randomisation is that all units have a non-zero probability of receiving the treatment, meaning the positivity assumption holds to ensure identifiable, unbiased treatment effects. Moreover, the controlled nature of the experimental design helps satisfy SUTVA. The controlled environment means no interference between participants, so one person's exposure does not affect another's outcomes. With these conditions for internally valid results met, causal effects are interpreted as Average Treatment Effects (ATE) (Druckman, 2022: 30):

$$\text{ATE} = \mathbb{E}[Y_t(u) - Y_c(u)] \tag{2}$$

The primary source of internal validity comes from randomisation, a condition difficult to achieve outside of controlled experiments. Angrist and Pischke (2010) argue that observational studies, such as natural experiments, can also exploit randomisation for causal inference. These studies use natural events and phenomena to generate 'as-good-as-randomly-assigned' treatments, but still require sophisticated statistical controls for possible confounders meaning exogeneity is not guaranteed. Notwithstanding issues of endogeneity, natural experiments are difficult to identify, and examples are especially localised meaning they have no advantage

[3]Stratified Random Assignment can help ensure heterogeneous characteristics are distributed across groups.

over laboratory experiments for increasing external validity. In the fake news literature, Cantarella, Fraccaroli and Volpe (2023) exploit a natural experiment through an instrumental variable of the proportion of Italian-to-German-speaking voters in each municipality to vary exposure to fake news. However, potential confounders of political and news preferences, or cultural differences, may affect consumption of fake news. With potential endogeneity, and SUTVA easily violated from spillover effects between groups, this example shows how a non-laboratory experiment brings multiple challenges for estimating causal effects of AI-generated articles on political attitudes. The final section outlines the cost of external validity and other trade-offs when using laboratory experiments.

Limitations of Laboratory Experiments

The primary limitations of laboratory experiments are interconnected issues of validity and realism. Firstly, external validity — where cause-effect relationships hold across people, treatments, and scenarios, i.e., does the causal relationship generalise? — is most prominent (Druckman, 2022: 61). Due to the laboratory's artificial environment the 'causal effect is always local, derived from a particular time, place, and research design' (Angrist and Pischke, 2010: 23). Whilst this issue is significant for policy-focused studies concerned with effect size, my research prioritises observing the effect direction, making generalisability less critical. Moreover, heterogenous covariates across nations such as political trust and media literacy limit comparisons outside of the UK, especially if these variables have varying moderation and mediation effects altering how individuals interact with AI-generated news. But, to maximise external validity, an Independent and Identically Distributed (IID) sample of UK-based voters will be used. Working with a partner such as YouGov will ensure a representative and accurately weighted sample of the UK, helping generalise results within UK settings (Aronow and Samii, 2016: 261; YouGov, 2024).[4] Secondly, a contributing factor to external validity is realism. Mundane realism considers whether the treatment and its setting represent the real world (Druckman, 2022: 51). Field experiments representing a randomised study conducted in a real-world setting help solve for the aspirational ecological validity of mundane realism (Gerber and Green, 2012: 10). But, field experiments struggle to control and measure complex mechanisms, and the 'high dimensionality' of interactions which determine the formation of political attitudes meaning important moderators can be missed (Kocher and Monteiro, 2016: 954). Instead, theoretically informed moderators can be included and isolated in an laboratory experiment to test interaction effects, and for use in later mediation analysis of causal mechanisms, giving the possible experiment design:

[4] Angrist and Pischke (2010) recommend multiple replicable studies to increase external validity, but this is outside of scope.

Table 1: AI-generated News Control vs Treatment conditions.

	AI-generated News Articles	
	True Articles	**False Articles**
Control	Labeled as AI	Labeled as AI
Treatment	No label	No label

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

Furthermore, experimental realism ensures participants authentically engage with the experiment despite the artificial laboratory environment. This predominantly focuses on internal validity affected by spillover effects violating SUTVA, such that treatments should be designed in an engaging and natural format representing how news is consumed (Druckman, 2022: 52). Moreover, a between-subjects design to separate groups and avoid interference — whilst avoiding issues of temporal stability and causal transience in a within-subjects design — is used (Holland, 1986: 948).

This essay has argued that a between-subjects randomised laboratory experiment is the most suitable research design for evaluating causal effects of AI-generated news on political attitudes. Laboratory experiments are shown as the most effective choice over qualitative or natural experiments due to their use of a large-N approach and guaranteed randomisation for strong internal validity despite the possible limitations of external validity and realism.

`Word Count: 1,999`

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.

Angrist, J.D. and Pischke, J.-S. (2010) 'The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics', *The Journal of Economic Perspectives*, 24(2), pp. 3–30.

Aronow, P.M. and Samii, C. (2016) 'Does Regression Produce Representative Estimates of Causal Effects?', *American Journal of Political Science*, 60(1), pp. 250–267.

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.

Brady, H.E. and Collier, D. (2010) *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Blue Ridge Summit: Rowman & Littlefield Publishers, Incorporated.

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*.

Druckman, J.N. (2022) *Experimental Thinking: A Primer on Social Science Experiments*. Cambridge: Cambridge University Press.

Druckman, J.N., Greene, D.P., Kuklinski, J.H. and Lupia, A. (eds) (2011) *Cambridge Handbook of Experimental Political Science*. Cambridge: Cambridge University Press.

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.

Gerber, A.S. and Green, D.P. (2012) *Field experiments: Design, analysis, and interpretation*. New York: Norton.

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

Goertz, G. and Mahoney, J. (2012) *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. 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.

Holland, P.W. (1986) 'Statistics and Causal Inference', *Journal of the American Statistical Association*, 81(396), pp. 945–960.

King, G., Keohane, R.O. and Verba, S. (1994) *Designing social inquiry scientific inference in qualitative research*. Princeton, N.J: Princeton University Press.

Kocher, M.A. and Monteiro, N.P. (2016) 'Lines of Demarcation: Causation, Design-Based Inference, and Historical Research', *Perspectives on Politics*, 14(4), pp. 952–975.

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).

Przeworski, A. (2009) 'Is the Science of Comparative Politics Possible?', in C. Boix and S.C. Stokes (eds) *The Oxford Handbook of Comparative Politics*. Oxford: Oxford University Press.

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.

Titiunik, R. (2015) 'Can Big Data Solve the Fundamental Problem of Causal Inference?', *PS: Political Science & Politics*, 48(1), pp. 75–79.

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.

YouGov (2024) 'Methodology | YouGov'. https://yougov.co.uk/about/panel-methodology.