

# University of Oxford: MPhil in Politics

## YouGov Survey Design

Edward Anders

### Research Project and Aims

Machine learning advancements to efficiently handle sequential data inputs and outputs have popularised the field of Artificial Intelligence (AI). 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. This research project is interested in AI's capacity to 'amplify social injustice, erode social stability, [...] customised mass manipulation, and pervasive surveillance' (Bengio *et al.*, 2024). These social and political risks of AI are often discussed anecdotally, but there remains little research nor evidence on what these risks look like. The UK Government's Department for Science, Technology & Innovation (2025) views 'manipulation and deception of populations' a significant threat to political systems and societies; but, the extent to which politically targeted generative AI can be used to distort, deceive, and direct an electorate remains unclear. Therefore, this project aims to answer:

*Does exposure to AI-generated political content increase affective polarisation?*

After running an initial pilot survey experiment with YouGov, it was found that AI-generated and AI-labelled content did not have statistically significant effect on either thermometer ratings nor trait- and emotional-based attitudes towards the respondents' out-group. However, certain subgroups showed clearly increased affective polarisation measures. To better understand the mechanisms behind these results, and in particular test the emotional and cognitive pathways suggested by the literature, I hope to run a second experiment testing respondents' assignments of particular emotional feelings towards the content in the AI-generated and AI-labelled conditions.

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

## Draft Survey Questions

These questions will be used to measure whether AI-generated and AI-labelled content triggers statistically significantly different emotional reactions than those from the control group.

## Post-Exposure Survey Questions

No.	Question	Options
<b>1</b>	<b>Strength of Political Identity</b>	
1.1	When people criticise <i>[your preferred party]</i> , it feels like a personal insult.	5-point Likert: Strongly disagree – Strongly agree
1.2	When I speak about <i>[your preferred party]</i> , I usually say “we” instead of “they.”	5-point Likert: Strongly disagree – Strongly agree
1.3	If <i>[your preferred party]</i> performs poorly in an election or poll, my day is ruined.	5-point Likert: Strongly disagree – Strongly agree
<b>2</b>	<b>Emotional Responses to Political Content</b>	
2.1	How angry did you feel while viewing this content?	Numeric scale: 0 (Not at all angry) – 10 (Extremely angry)
2.2	How anxious or worried did you feel about the issue presented?	Numeric scale: 0 (Not at all anxious) – 10 (Extremely anxious)
2.3	How disgusted did you feel while viewing this content?	Numeric scale: 0 (Not at all disgusted) – 10 (Extremely disgusted)
2.4	To what extent do you trust the information presented in the video/article?	5-point Likert: Do not trust at all – Completely trust
<b>3</b>	<b>Partisan Schadenfreude</b>	
3.1	I am happy when negative things happen to <i>[the opposing party or its supporters]</i> .	5-point Likert: Strongly disagree – Strongly agree
3.2	I would be pleased by the misfortune that happened to a politician from <i>[opposing party]</i> .	5-point Likert: Strongly disagree – Strongly agree

## References

Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Darrell, T., Harari, Y.N., Zhang, Y.-Q., Xue, L., Shalev-Shwartz, S., Hadfield, G., Clune, J., Maharaj, T., Hutter, F., Baydin, A.G., McIlraith, S., Gao, Q., Acharya, A., Krueger, D., Dragan, A., Torr, P., Russell, S., Kahneman, D., Brauner, J. and Mindermann, S. (2024) ‘Managing extreme AI risks amid rapid progress’, *Science*, 384(6698), pp. 842–845.

Department for Science, Technology & Innovation (2025) ‘Safety and security risks of generative artificial intelligence to 2025 (Annex B)’, *GOV.UK*. <https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper/safety-and-security-risks-of-generative-artificial-intelligence-to-2025-annex-b>.