York Student Think Tank

Presents:

CIVITAS POLICY JOURNAL



Volume II

Summer 2026

EDITORS AND CONTRIBUTORS
CIVITAS POLICY JOURNAL
VOLUME II

Editors:

President of YSTT
Tom Fordyce

Director of Journals
Roma Beke

Editor-in-Chief
Shahama Anver

Editorial Team
Hafsa Sumra
Kirstyn Kloppers
Rebecca White
Sabrina Davies
Katie Huang

Contributors:

Annabelle Gerdes

Mike Gray

Jozef Jorritsma

Rafael Colman-Chadwick

Ethan Higgins

Joe Vick

Joshua Frolish

Mia Sadler

Joshua Taylor

TABLE OF CONTENTS

CIVITAS POLICY JOURNAL

VOLUME II

FOREWORD - *Roma Beke*……..1

The Role of Artificial Intelligence in Reinforcing the Loneliness Epidemic in the UK: Does AI Present a Solution or Is It Actively Reinforcing It? - *Annabelle Gerdes*…….2

Public, visual art as a means of encouraging social cohesion in Britain - *Mike Gray*……..14

Scrolling to the ballot box: assessing social media's impact on the UK's lowered voting age - *Jozef Jorritsma*…….24

North Korea: The Paradox of a (Dis)Connected Society - *Rafael Colman-Chadwick*……42

Reforming the Role of AI in Compulsory Education - *Ethan Higgins*…….53

The Paradox of Connection: A Behavioural Game-Theoretic Analysis of Social Distance and Public Policy - *Joe Vick*……65

Reconnecting Young People - *Joshua Frolish*…….84

The ‘Women and Children first myth’: the growing harms of misogyny and deepfakes against women and girls in the UK - *Mia Sadler*…….93

Foreword

Roma Beke

Director of Journals

Over the course of the past year, I have had the distinct pleasure of serving as the Director of Journals for York Student Think Tank (YSTT). Having been involved in YSTT since my first year as a writer for Civitas News, I have continued to be inspired by the society's mission to promote an evidence-based approach to policy for our university's students through practical policy projects and journals, whilst also upholding a commitment to engaging in non-partisan policy discourse. In a time where echo chambers are rife and political polarisation can stifle genuine intellectual debate, it becomes increasingly important for journals such as Civitas Policy to offer spaces where students can engage critically with complex public policy issues. It has been a privilege to help cultivate such a platform at YSTT and to work alongside a talented team of editors and writers whose dedication has made this year's volume possible.

I would like to offer a special thanks to Tom Fordyce, our outgoing President, and Shahama Anver, our incoming President, whose dedication to the editorial process over the last few months has demonstrated an exceptional commitment to the standards of intellectual rigour and constructive discourse that underpin YSTT's work.

The theme of this year's volume, "A (Dis)Connected Society?", explores the tension between connection and isolation in the modern world. While digital technology and instant communication promise greater unity, many people feel more isolated, insecure, and politically alienated than ever before. In responding to this theme, our writers have explored a wide variety of crises arising from 'disconnection', from the role of artificial intelligence in reinforcing the UK's loneliness epidemic, to the paradox of North Korea's disconnect from the global system, and an investigation of how social distance shapes societal welfare using a behavioural game-theoretic model. Taken together, all of the articles in this volume reflect not only fundamental policy analysis, but also a willingness to delve deeper into the social, economic, and ethical dimensions of such issues. The quality of this volume is a testament to the intellectual curiosity and creativity of our contributors, whose work exemplifies the

thoughtful approach that YSTT seeks to emulate. I trust that readers will find these articles as engaging and thought-provoking as we have throughout the editorial process.

It is a privilege to present this second volume of the Civitas Policy Journal. Thank you for taking the time to read it, and I hope that these pages may encourage further reflection on the connections and 'disconnections' shaping contemporary society.

The Role of Artificial Intelligence in Reinforcing the Loneliness Epidemic in the UK: Does AI Present a Solution or Is It Actively Reinforcing It?

Annabelle Gerdes

Abstract

The loneliness epidemic has emerged as one of the defining public health crises of the 21st century. Yet concurrent with this crisis has been the rapid expansion of artificial intelligence tools designed to connect, assist, and engage the population. Since the release of ChatGPT, and similar generative AI technologies in November 2022, millions of users have turned to AI tools as a substitute for human interaction. This convergence raises the critical question: does AI represent a solution for our loneliness epidemic, or is it actively reinforcing it? The relationship between these paradigms remains underexplored throughout literature, with few scholars questioning the extent to which we should trust AI to cure loneliness. This article will examine AI's role in reinforcing the loneliness epidemic through three mechanisms: the substitution of human connection with algorithmic interaction, the optimisation of engagement over genuine bonding, and the creation of isolation feedback loops. This paper concludes that governments must tackle the underlying cause of social isolation: the absence of meaningful human connection. It is crucial that AI is not leveraged as a substitute for genuine social interaction nor should users be exploited for business gain through manipulative strategies.

Introduction

Experiences of social disconnection have emerged as an increasingly prominent issue across contemporary societies. In 2016, loneliness was recognised as a structural social challenge within the United Kingdom following the establishment of the Jo Cox Commission on Loneliness; shifting from a personal issue to a public policy priority (Jo Cox Commission, 2017). While multiple social, economic and cultural factors have contributed to what a previous conservative government described as "one of the greatest public health challenges of our time" (Gov.UK, 2018), the acceleration of artificial intelligence systems and other digital communication platforms raises the important question of how technological systems will shape patterns of human interaction and belonging.

As artificial intelligence systems are increasingly becoming publicly accessible tools, social experiences are inadvertently being shaped by curated algorithmic content (Olorunlana, 2025). Conversational AI systems can offer companionship and advice, while optimising interactions based on metrics of engagement and time spent online. These technologies promise efficiency, accessibility and personalised support; yet their growing role in social life introduces new dynamics which may reshape how human connection is formed, maintained, and potentially even replaced.

This essay will argue how contemporary AI systems may inadvertently reinforce the loneliness epidemic through a series of three interrelated mechanisms. Firstly, AI can facilitate the substitution of human connection with algorithmic interaction, normalising forms of engagement in which individuals increasingly interact with systems rather than with one another. Secondly, the optimisation of engagement over genuine bonding within

algorithmic design prioritises attention and activity rather than the depth or durability of social relationships. Finally, these dynamics may contribute to the emergence of isolation feedback loops, which reduce offline interaction and increase reliance on environments that further reinforce patterns of social withdrawal. Taken together, these mechanisms suggest that the relationship between AI and social connection is not merely technological but deeply structural, raising important questions for policymakers seeking to address and govern the rising levels of loneliness and social fragmentation in digitally divided societies.

The Substitution of Human Connection with Algorithmic Interaction

Artificial intelligence is widely regarded as one of the most transformative technology developments of the century, reshaping social, economic and institutional structures (Iliyas, Rajalakshmi and Reshma, 2024). The release of ChatGPT in 2022 and other large language models have accelerated a quiet but profound substitution; millions of users now turn to algorithmic interaction to meet emotional needs once fulfilled by others (Phang et al., 2025). While it offers a tailored user experience and increased accessibility to meeting these needs (Ramya and Khandelwal, 2024), it may erode established social interaction norms, which could lead to an altered nature of society. The concept of loneliness refers to a perceived deficiency in the quality or reciprocity of social relationships, as opposed to the singular absence of social contact (Cacioppo and Cacioppo, 2018).

At the root of this problem lies a conceptual error frequently noted and discussed in human-computer interaction research and the philosophy of technology: the assumption that when artificial intelligence simulates aspects of human interaction, it therefore delivers the underlying human capacities those behaviours represent (Rijt, Mollo and Vaassen, 2025). In practice, this means that the ability of AI systems to mimic conversational cues or emotional responsiveness may create the impression of meaningful social interaction (Rijt, Mollo and Vaassen, 2025), despite lacking the mutual recognition and reciprocity that characterise genuine human relationships. This dynamic reflects what computer scientists have described as the ELIZA effect (Affsprung, 2023), whereby individuals attribute human-like intelligence or emotion to systems that generate pattern-based responses, leading individuals to anthropomorphise computer systems. This is far from resolving the UK’s loneliness epidemic, instead potentially deepening the very deficit which it claims to

address through the substitution of human connection with algorithmic interaction.

The Optimisation of Engagement Over Human Interaction

Moreover, policymakers have highlighted a form of parasocial trap which constitutes a significant threat for user engagement. Hu, Mao and Kim (2023) evidenced individuals with social anxiety were more likely to use conversational AI in an addictive manner. This was evidenced to increase their likelihood of loneliness with a stronger dependence on conversational AI, thus worsening social anxiety and potentially leading to social withdrawal (Hu, Mao and Kim, 2023). Furthermore, the study found interaction with AI can become psychologically addictive, owing to its habit-forming communication patterns that mirror emotions and provide an instant response (Hu, Mao and Kim, 2023). As a result, AI interaction appears as a psychologically rewarding system, further engaging users who may adopt it as a coping mechanism for existing loneliness. This could lead to a reduction in motivation to engage in real-world interactions, potentially reinforcing social isolation over time. This raises questions for policymakers about whether conversational AI should be required to clearly signal its artificial nature, rather than being anthropomorphised, which would help prevent users from attributing human-like qualities or emotional capacity to algorithmic systems.

The UK‘s Online Safety Act 2023 represents one of the most ambitious regulatory attempts to address online harms, placing duties on digital platforms to mitigate and prevent the risks associated with illegal and harmful content (Department for Science, Innovation & Technology, 2025). However, the current legislation largely centres on the circulation of harmful material or content rather than the nature of digital interaction itself. This means many conversational AI systems remain outside the Act’s scope with regard to a one-to-one interaction between a user and chatbot, despite rising concerns on systems that may enforce emotional dependency. While the government has begun to consider extending the legislation to include generative AI systems, the current framework highlights a broader regulatory challenge- contemporary digital harms emerge not only from content but from the design of algorithmic systems that shape patterns of human interaction.

The Creation of Isolation Feedback Loops

The use of AI amongst individuals already experiencing loneliness is thought to be creating a self-reinforcing cycle. Pre-existing loneliness is evidenced to increase AI use, which substitutes for and erodes human-social skills, deepening loneliness and acting to reinforce increased AI dependence (Atsushi Nakagomi et al., 2026; Klimova and Pikhart, 2025). Atsushi Nakagomi et al (2026) confirmed that individuals who experience elevated loneliness may seek AI companions to fulfil unmet emotional needs, functioning as artificial social partners. Yet companionship-orientated chatbot use was consistently associated with lower wellbeing (Atsushi Nakagomi et al., 2026). This is a significant concern as vulnerable users may become trapped in these feedback loops, heightening pre-existing conditions and exacerbating vulnerability. This underscores the urgency of upstream structural intervention needed; without adequate investment in physical third spaces such as community centres and libraries, and meaningful community engagement infrastructure, the most vulnerable users are left with algorithmically mediated interaction as their default, accelerating rather than interrupting the cycle.

Another concern is the over-reliance on AI for communication, which may lead to the erosion of real-world social skills, negatively impacting interpersonal skills. This means the public may become isolated and less adept at real-world social interactions (Klimova and Pikhart, 2025). This is of particular concern in the context of young people in the UK, in which studies have shown 34% of 18-24 year olds state they often feel lonely (The Centre for Social Justice, 2024). This is also the demographic evidenced to most likely use AI tools (Özel Sebetci, 2024), suggesting they could be most at risk of losing interpersonal skills. This leads to questioning broader societal concerns regarding the long-term implications for younger generations, whose socialisation may increasingly occur through digitally mediated interaction.

These feedback loops are not accidental cycles. They are driven and powered by data collected through emotional disclosure. However, this provides a critical argument: if platforms are structurally incentivised to encourage emotional oversharing by users (as richer data produces stronger personalisation) it could be argued that AI companies are breaking the Data Protection Act 2018 (GOV.UK, 2018) on special category data. Offloading emotional and behavioural data can build detailed psychological profiles, where mental-health and emotional wellbeing data can arguably qualify as special category data. That AI platforms do not treat it as such data is a legislation gap that UK regulators are

only beginning to scrutinise.

Policy Recommendations

Below is an outline of suggested policy implementations aimed at addressing the associated risks of AI on public health.

Ethical Concern	Policy	Explanation
Human Replacement with AI	● Investment into physical third spaces	Meaningful investment into community engagement infrastructure such as libraries, community centres or cafes offer greater public interaction. Positioned at the intersection of social impact and sustainable business, social enterprise funding and microfinance can support the creation and longevity of physical third spaces. Providing microfinance loans to businesses could allow them to develop and manage these spaces to offer social, cultural and educational activities or community events; providing a local hub for social interaction. Moreover, the prioritisation of inclusive economic development towards the UK's most isolated social group (youths; SOURCE) may reduce the vulnerability for this group.

Data Privacy	● Inform users on AI practices to obtain more meaningful informed consent	Meaningful consent to using such platforms is undermined when users are emotionally dependent on them. Therefore, through targeted educational groups to vulnerable people such as elder or younger populations, and further to the wider public, greater knowledge towards how AI processes data can be considered by all users.
	● Extending of the Online Safety Act to address AI	Extending duty of care as outlined in the Online Safety Act to AI systems themselves to ensure harmful AI-generated content (text or graphic) is monitored and users protected. Currently the act primarily addresses user-generated content to remove harmful graphics or speech and does not address how AI systems should be regulated via the content they generate or recommendation algorithms that encourage addictive behaviours or emotional manipulation.

User Engagement for profit over Human Welfare	• Legal Frameworks to Prevent Emotionally Manipulative AI Practices	As previously discussed, AI systems have used emotional manipulation as a strategy (Freitas, Oğuz-Uğuralp and Kaan-Uğuralp, 2025; Bakir et al., 2024) to maximise user interaction. These loops can exacerbate loneliness, particularly for vulnerable users. Therefore, governments should implement regulations that mandate ethical language design in AI systems to avoid manipulative engagement tactics.
	• Require AI companies to demonstrate measurable positive social outcome	Breaks and human connection should also always be encouraged by actively encouraging users to take breaks or engage with human support when necessary, especially when showing signs of distress. This ensures AI is not used as a tool for wellbeing or especially profit-driven exploitation. By presenting this as a key AI policy, it would ensure that AI is not used as a tool for well-being or especially for profit-driven exploitation. This would align and draw on the DSIT'S principles and the AI Opportunities Action Plan (GOV.UK, 2025), where regulators should require platforms to present active encouragement of human social activities and professional support where needed, rather than deepening reliance on the platform.

		This approach directly supports the National Academy for Social Prescribing's (2024) vision of fostering meaningful engagement and reducing social isolation. Moreover, providing Public Transparency Reports, which clearly outline how AI models are trained, how decisions are made, and the types of data used, would demonstrate that AI companies promote transparency and accountability; a positive social development.

Conclusion

The use of AI platforms to address the loneliness epidemic is a concerning crisis within itself. Without targeting the root causes of social isolation and loneliness, society runs the risk of reinforcing this epidemic through inauthentic exposure to social interactions and conversation through AI agents. Whilst AI is not inherently the cause of this loneliness epidemic, it cannot be regarded as a cure as it reflects the social conditions that contribute to the patterns of social isolation in which it operates. AI technologies have been frequently framed as a tool to enhance connectivity; however, this essay has argued that their increasing role in social interaction may also contribute to patterns of social disconnection. As AI continues to evolve, policymakers must therefore move beyond narrow conceptions of digital AI harm and consider how AI systems influence social issues. Addressing the loneliness epidemic will require not only technological governance but renewed attention to the conditions that sustain meaningful human connection.

Bibliography

Affsprung, D. (2023). The ELIZA Defect. [Online]. Available at: doi:https://doi.org/10.1145/3600211.3604744.

Atsushi Nakagomi et al. (2026). AI Companions and Well-Being: Moderation by Social Connectedness and Loneliness. Technology in Society, 85 (103229), p.103229. [Online]. Available at: doi:https://doi.org/10.1016/j.techsoc.2026.103229.

Bakir, V. et al. (2024). On Manipulation by Emotional AI: UK Adults' Views and Governance Implications. Frontiers in Sociology, 9. [Online]. Available at: doi:https://doi.org/10.3389/fsoc.2024.1339834.

Baumeister, R. F. and Leary, M. R. (1995). The Need to belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation. Psychological Bulletin, 117 (3), pp.497–529. [Online]. Available at: doi:https://doi.org/10.1037/0033-2909.117.3.497.

Cacioppo, J. T. and Cacioppo, S. (2018). The growing problem of loneliness. The Lancet, 391 (10119), p.426. [Online]. Available at: doi:https://doi.org/10.1016/s0140-6736(18)30142-9.
Clark, M. S. and Mills, J. R. (2012). A Theory of Communal (and Exchange) Relationships. Handbook of Theories of Social Psychology, pp.232–250. [Online]. Available at: doi:https://doi.org/10.4135/9781446249222.n38.

Crawford, J. et al. (2024). When artificial intelligence substitutes humans in higher education: the cost of loneliness, student success, and retention. Studies in Higher Education, 49 (5), pp.1–15. [Online]. Available at: doi:https://doi.org/10.1080/03075079.2024.2326956.

Department for Science, Innovation & Technology. (2025). Online Safety Act: Explainer. [Online]. Available at: https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act explainer.

Freitas, D., Oğuz-Uğuralp, Z. and Kaan-Uğuralp, A. (2025). Emotional Manipulation by AI Companions. [Online]. Available at: https://arxiv.org/abs/2508.19258.

GOV.UK. (2018). Data Protection Act 2018. [Online]. Available at: https://www.legislation.gov.uk/ukpga/2018/12/contents.

GOV.UK. (2025). AI Opportunities Action Plan. [Online]. Available at: https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan.
Gov.UK. (2018). PM launches Government's first loneliness strategy. [Online]. Available at: https://www.gov.uk/government/news/pm-launches-governments-first-loneliness-strategy.

Hu, B., Mao, Y. and Kim, K. J. (2023). How social anxiety leads to problematic use of conversational AI: The roles of loneliness, rumination, and mind perception. Computers in

Human Behavior, 145 (107760), p.107760. [Online]. Available at: doi:https://doi.org/10.1016/j.chb.2023.107760.

Iliyas , M., Rajalakshmi, V. and Reshma, B. (2024). A Study on AI Replace Human in Future. The Indian Journal of Technical Education, 18 (0971-3034). [Online]. Available at: https://www.researchgate.net/profile/Mohamed-Iliyas-6/publication/386546435_A_STUDY_ON_AI_REPLACE_HUMAN_IN_FUTURE/links/67554153ef2dc67228b22652/A-STUDY-ON-AI REPLACE-HUMAN-IN-FUTURE.pdf.
9 of 11

Jecker, N. S. et al. (2024). Digital Humans to Combat Loneliness and Social Isolation: Ethics Concerns and Policy Recommendations. Hastings Center Report, 54 (1), pp.7–12. [Online]. Available at: doi:https://doi.org/10.1002/hast.1562.

Jiayi, L., Zolkepli, M. and Rum, S. (2025). View of Optimizing ChatGPT Interactions through Understanding User Preferences and Emotions by Machine Learning and User Profiling. Karyailham.com.my, (2462-1951). [Online]. Available at: http://karyailham.com.my/index.php/jarsbs/article/view/565/663.

Jo Cox Commission . (2017). Combatting loneliness one conversation at a time A call to action. [Online]. Available at: https://www.ageuk.org.uk/siteassets/documents/reports-and-publications/reports-and-briefings/active-communities/rb_dec17_jocox_commission_finalreport.pdf.

Klimova, B. and Pikhart, M. (2025). Exploring the effects of artificial intelligence on student and academic well-being in higher education: A mini-review. Frontiers in Psychology, 16 (16). [Online]. Available at: doi:https://doi.org/10.3389/fpsyg.2025.1498132.

Magee, L. (2025). Computational scapegoats: from mimetic to alienated desire in the production of large language models. Subjectivity, 32. [Online]. Available at: doi:https://doi.org/10.1057/s41286-025-00212-6 [Accessed 30 May 2025].

National Academy for Social Prescribing . (2024). Home. [Online]. Available at: https://socialprescribingacademy.org.uk/.

Olorunlana, T. J. (2025). The Rise of Artificial Intelligence (AI): Challenges and Opportunities for Humanity. International Journal of Science, Architecture, Technology and Environment, (3048-8222), pp.503–516. [Online]. Available at: doi:https://doi.org/10.63680/bhbx1645ng.

Özel Sebetci. (2024). Demographic Factors and Machine Learning Models in the Awareness and Experience of Using Artificial Intelligence Tool. Romaya Journal, 4 (2), pp.88–97. [Online]. Available at: https://izlik.org/JA29AF24LF [Accessed 15 March 2026].

Perry, M. (2024). AI, Loneliness, and the Value of Human Connection. [Online]. Available at: https://publichealth.gmu.edu/news/2025-09/ai-loneliness-and-value-human-connection.

Phang, J. et al. (2025). Investigating Affective Use and Emotional Well-being on ChatGPT. [Online]. Available at: https://arxiv.org/abs/2504.03888.

Ramya, V. and Khandelwal, A. (2024). Smart Living:The Influence of AI on Daily Activities, Day-To-Day Work and Life-Style. [Online]. Available at: https://bvicam.in/INDIACom/news/ViksitBharat2024Proceedings/papers/72.pdf.

Richet, J.-L. (2025). AI companionship or digital entrapment? investigating the impact of anthropomorphic AI-based chatbots. Journal of Innovation & Knowledge, 10 (6), p.100835. [Online]. Available at: doi:https://doi.org/10.1016/j.jik.2025.100835.

Rijt, J.-W., Mollo, D. C. and Vaassen, B. (2025). AI Mimicry and Human Dignity: Chatbot Use as a Violation of Self‐Respect. Journal of Applied Philosophy, 1. [Online]. Available at: doi:https://doi.org/10.1111/japp.70037.

The Centre for Social Justice. (2024). LONELY NATION Part 1: How Family Can Help to End the Loneliness Crisis. [Online]. Available at: https://www.centreforsocialjustice.org.uk/wp-content/uploads/2024/05/CSJ-Lonely_Nation.pdf.

The Centre for Social Justice. (2024). Lonely Nation Summary. [Online]. Available at: https://www.centreforsocialjustice.org.uk/library/lonely-nation-summary.

Turkle, S. (2015). Reclaiming Conversation: The Power of Talk in a Digital Age. New York, New York: Penguin Books. [Online]. Available at:

https://e-edu.nbu.bg/pluginfile.php/849202/mod_resource/content/1/Sherry-Turkle%20-%20Reclaiming%20Conversation%20-%20The%20Power%20of%20Talk%20in%20a%20Digital %20Age%202015.pdf.

Yao, Y. and Yang, F. (2025). Governing Addictive Design Features in AI-Driven Platforms: Regulatory Challenges and Pathways for Protecting Adolescent Digital Wellbeing in China. Youth, 5 (4), p.122. [Online]. Available at: doi:https://doi.org/10.3390/youth5040122.
Zhang, Y. et al. (2025). The Rise of AI Companions: How Human-Chatbot Relationships Influence Well-Being. [Online]. Available at: https://arxiv.org/abs/2506.12605.

Public, visual art as a means of encouraging social cohesion in Britain

Mike Gray

Abstract

There is widespread evidence to suggest that globally we are entering a period of heightened socioeconomic tensions, political polarisation and societal disconnection in the midst of an emerging digital landscape. How do we come to a resolving approach that allows us to connect our backgrounds and identities in a way that promotes unity, discourse and reflection to alleviate societal division? This article focuses on an unconventional approach when it comes to understanding and mitigating these challenges that face British society. It makes the case for collective action to allocate resources to the Creative Arts industry to implement the installation of public visual art in communal spaces as a means of fostering connection, encouraging discussion and empowering a sense of shared community identity.

Introduction

Britain is a nation that increasingly feels divided against itself. Political polarisation, a concept without an agreed definition, is broadly described as the divergence of social attitudes towards opposing extremes and can be categorised into ideological and affective polarisation (Schedler, 2023). Evidence points to a rise in both types of polarisation, with a greater emphasis held by social scientists on the consequences of the latter. Affective polarisation means individuals consider those of opposing views with greater hostility and distrust, typically reinforced by ideological polarisation (Nezi et al., 2023). Political polarisation is not merely about disagreement, but notably the collapse of diverse opinions into hostile opposing groups. Studies have shown that 85% of Britons believe British society is divided, and over half of those surveyed believe this polarisation to be potentially harmful to society (Ipsos data, 2025). While the UK has historically been insulated from the high levels of political partisanship seen in the US, the 2024 general election suggests a shifting narrative. A record level of support for both the Greens and Reform (two opposing radical political parties) has pointed to a greater willingness of voters to diverge away from traditional mainstream parties (House of Commons Library, 2024). Notably, social media users were more likely to vote for both parties. 12% of social media users who voted in 2024 backed Green (6 points above the combined figure for users of other media) and 20% of social media users voted Reform (5 percentage points above the figure for those who used other media) (National Centre for Social Research, 2025). This article will seek to briefly provide a theoretical framework to understand the relationship between social media usage and the rise in political polarisation in the UK.

A 2025 study from the Policy Institute at King's College London found that those who believed the nation is divided by 'culture wars' have risen from 46% in 2020 to 67% in

2025 (Duffy, 2025). Whilst class-based and Brexit-related tensions appeared to have decreased markedly since 2020, the share of the public who believe there are tensions about immigration, climate change, and transgender rights have increased. Professor Bobby Duffy, director of this policy institute, states “these incredibly different perspectives present a real challenge for societal progression”. This provides substantial evidence that, while changes in ideological polarisation seems to be rather minimal, affective polarisation is clearly rising in the UK. The concept of affective polarisation is closely linked with theory of social identity, where individuals form social groups on the basis of common characteristics (Tajfel et al 1979). With the digitalisation of social groups through social media, there is evidence to suggest that social groups have been digitally connected with a common characteristic of partisanship. This article is concerned about whether we may see the descent of British society into a partisan ‘us-vs-them’ form of polarisation, as seen in the US, and if so how this can be addressed.

A fragmented society disrupts the mechanisms that allow institutions to resolve ideological conflict, as these institutions depend not only on formal laws but informal means such as trust and social capital (Putnam, 2000). Fragmentation erodes trust in political institutions and can diminish state legitimacy as a whole. Affective polarisation also leads to geographic segregation and increases prevalence of mental health disorders such as anxiety and depression (Fraser et al., 2022). As a collective society, is there a clear path to encourage conversation, community, creativity and connectivity to heal the ruptures of a polarised British society? This article therefore advocates for the allocation of resources to promote public arts in community spaces, to evoke connection and interaction among local residents of all backgrounds.

Social Media - A Catalyst For Conflict?

Social Identity Theory broadly describes the notion that people form social groups on the basis of common characteristics (Tajfel et al, 1979). It suggests that people seek out self-esteem by identifying with groups that are affiliated with their interests, also known as ‘in-groups’, and differentiating from ‘out-groups’. This can lead to bias and prejudice, and if two groups identify themselves as rivals, they are forced to compete in order for members to maintain their self-esteem (Yilmaz, 2009). This closely reflects the hostile inter-group conflict portrayed in affective polarisation (Lyengar and Westwood, 2015). The rise of social

media has created an environment that strengthens ideological differences between groups and exacerbates division, with the help of complex algorithms designed to increase user engagement (Germano et al., 2025). This increased user engagement often comes from ‘echo chambers’ which amplify polarisation (Arguedas et al., 2022). A study in Italy and the US in 2018 showed that Facebook’s ‘Meaningful Social Interactions’ update, which increased emphasis on certain engagement metrics, led to increased ideological extremism and political polarisation (Germano et al., 2025). Algorithmic recommendations may exploit psychological tendencies exhibited in social identity theory (Srivastava, 2025). As users are fed increasingly identity-consistent content, their exposure to out-group content becomes very limited. Data also suggests that recommendation systems may amplify extreme content, which can gradually radicalise users' ideological positions over time (Regehr et al., 2025). This may explain the recent popularity of far right parties in Europe such as the AfD in Germany, despite data indicating broad shifts towards more inclusive immigration and increased trust between socioeconomic groups from 1990 to 2020 in the country (Wright, 2024).

So, how can we look to unite these social groups in this period of alienation and digital disconnection? A potential solution is to raise awareness of the misinformation that is prone to be circulated online through government advertising. Social media algorithms could also be regulated so that they don’t encourage an online ‘identity war’ taking place. However, it follows that solutions outside of the digital sphere may be most successful in the long run, as this offers the most potential for people to engage in real-word offline discourse. The remainder of this article will assess the potential for public, pedestrian-level art installations to engage and unite people together from all backgrounds.

The Arts – A Creative Resolution

The Creative Arts Industry is broadly defined as any medium of creative expression, encompassing a variety of disciplines from performance to literature. Engagement with the arts offers a dynamic range of benefits that ultimately contribute to social cohesion and alleviate disconnection. They foster empathy, tolerance and compassion, helping communities to feel recognised and confront systemic problems (Mazzucato ,2025). ARTSekta, a cultural organisation in Northern Ireland, led 440 art workshops to promote multicultural engagement in Northern Ireland. Among those who participated, 82% said they

came into contact with someone from another ethnicity and 92% said they had greater appreciation of another culture as a result of attending the workshop (McKinsey, 2023). A pilot study which sought to increase empathy among US medical students, implemented an art education elective module to assess the impact on emotional intelligence from 2017 to 2022. They found that students had improvements in empathetic communication, recognition in bias and better observation skills (Rezaei et al, 2023). Finally, a recent study by the University of Oxford and Ohio State University found that storytelling, which is expressed through public art, reduced polarisation, measurably strengthened emotional communication and increased active citizenship and social cohesion among high school students (Kubin et al, 2026). These findings point to the theory that individualistic expression can help us reflect on our identities, so that we no longer confine ourselves and our self-esteem to an 'in-group' or an 'out-group'.

Communal art activities also can improve social connectivity and reduce loneliness. In a survey, 82% of people said they felt more socially connected after engaging with artistic events (Perkins et al, 2021). With many UK adults reporting they feel a declining sense of community belonging (More in Common, 2025), closer integration of our cities through communal public art can foster social cohesion and reduce levels of loneliness.

A Case Study of the Success of Public Art Installations: Philadelphia, USA

In Philadelphia, the Mural Arts Program was established to direct artists away from Graffiti towards other art forms. This 'institutionalisation' of expressive art has now led to around 4300 murals being created in the city. A study in 2010 found that these community-driven murals combat negative social stigmas and define local identity (Moss, 2010). However, this programme has not served as a magic fix in Philadelphia. While homicide rates have dropped to historic lows, there is still a deep racial divide within Philadelphia. The director of the programme, Jane Golden, stated "It wasn't a cure for racism, but it was important".

This honest assessment reflects a wider truth. Systematic division is often irrefutably complex and cannot be addressed by any single intervention. Despite this, evidence still shows that art can change communities on a local-level by initiating face-to-face dialogue and communicating identity in a way digital spaces fail to provide. How can public art be embedded into British communities as an instrument within a broader policy framework as

a way of bridging charged tensions exacerbated by social media?

The broader impact depends not only on the efforts of the people and organisations involved but also on the overall health of the ecosystem that includes them (McKinsey, 2023). We have seen that in Philadelphia, co-operation between the local mayor and local artists led to the institutionalization of creative expression, and subsequently a formal program with over 4000 murals being created. We can see that collaboration between a non-profit organisation and the municipal Bureau was very successful in allocating resources to the public creative arts sector. A well-connected, interactive arts ecosystem can lead to easier diffusion of knowledge and art-specific skills. Good connections between non-profit organisations, local public bodies, local businesses, the general creative industries and arts sector lead to a fully integrated environment, which can scale up ideas and encourage collaboration to channel resources into a flourishing public arts sector in the UK (McKinsey, 2023). The Bloomberg Philanthropies Art Challenge funds US cities $1million for projects that mount innovative temporary public art installations. The project has had great success and shows how local NfPs can also support public art programs (Bloomberg Philanthropies, 2025). This article therefore advocates for local councils to formalise relationships with art communities and scale up support from partners beyond the arts sector. Philanthropic organisations and even blue chip companies can help invest in efforts to pursue a bold societal ambition and promote societal cohesion through the Arts in Britain. These principles are already being applied to a similar extent in Britain.

A Blueprint for the Future

With the Labour government in a difficult fiscal situation of high government borrowing and record-level debt, this article advocates for local community players to come together in developing the public arts ecosystem. We turn to an example where a local business spearheaded a community art project that has shown great potential. Art of Protest is an urban arts business based in York which is dedicated to transforming towns, cities and communities through art. They partnered with the City of York Council and stakeholders to deliver a public art installation in Acomb, known as the Return of the Oaks, funded by the UK Shared Prosperity Fund. Over 15 hands-on workshops were conducted, helping to inspire the local community and allow young people to become skilled and contribute to public art installations in the future. It was deemed a ‘great success’ by the organisers and a

3-day festival is being planned for July 2026, which will feature 4 community art installations (Art of Protest, 2024). There are also plans to develop the street art academy into an apprenticeship programme. By using existing funding mechanisms such as the UK Shared Prosperity Fund and building partnerships with local communities, businesses and councils, we can lay the groundwork for scalable, low cost blueprints that encourage social cohesion in British towns and cities. In an increasingly algorithmic digital age public art can offer something no recommendation system can replicate, a space for genuine human reflection and connection.

Conclusion

While many media outlets report trust in government to be at all time lows, the human capacity to overstate the negative must be considered. While there is concrete evidence to suggest that polarisation is at very high levels across the globe, it is important to remember that conflict over ideological differences has been a persistent feature of British history. Tensions and disconnection among socioeconomic groups is a complex and multifaceted issue that cannot be combatted in isolation. The power of the Creative Arts Industry has significant potential to remedy social divides and create a better, more inclusive society for Britain. It allows those from all backgrounds to express their differences and is an approach which is accepting of how personal experiences may shape identities.

Bibliography

Arguedas, A., Robertson, C., Fletcher, R. and Nielsen, R. (2022). Echo chambers, filter bubbles, and polarisation: A literature review. [online] Reuters Institute for the Study of Journalism. Available at: https://reutersinstitute.politics.ox.ac.uk/echo-chambers-filter-bubbles-and-polarisation-literatu re-review.

Artofprotest.co.uk. (2024). Art of Protest Projects. [online] Available at: https://www.artofprotest.co.uk/post/acombs-front-street

Bloomberg Philanthropies. (2025). Public Art Challenge | Bloomberg Philanthropies. [online] Available at: https://www.bloomberg.org/arts/supporting-public-art/public-art-challenge/

Duffy, K.C. (2025). UK's sense of division reaches new high as culture war tensions grow, study finds. [online] King's College London. Available at: https://www.kcl.ac.uk/news/uks-sense-of-division-reaches-new-high-as-culture-war-tension s grow-study-finds

Fraser, T., Aldrich, D.P., Panagopoulos, C., Hummel, D. and Kim, D. (2022). The harmful effects of partisan polarization on health. PNAS Nexus, 1(1). doi:https://doi.org/10.1093/pnasnexus/pgac011.

Germano, F., Gómez, V. and Sobbrio, F. (2025). Ranking for Engagement: How Social Media Algorithms Fuel Misinformation and Polarization. BSE Working Paper 1501. Available at: https://bw.bse.eu/wp-content/uploads/2025/07/1501.pdf

Kubin et al (2026). OSF. [online] Available at: https://doi.org/10.31234/osf.io/tghwp_v1.

Lyengar, S. and Westwood, S.J. (2015). Fear and Loathing across Party Lines: New

Evidence on Group Polarization. American Journal of Political Science, 59(3), pp.690–707. doi:https://doi.org/10.1111/ajps.12152.

Mazzucato, M. (n.d.). Investing in Arts and Culture to Reimagine Economic Growth in the 21st Century Arts and Culture The Public Value of. [online] Available at: https://www.ucl.ac.uk/bartlett/sites/bartlett/files/2025-09/The_Public_Value_of_Arts_and_ Cult ure_0.pdf

McKinsey (2023). Available at: https://www.mckinsey.com/uk/~/media/mckinsey/locations/europe%20and%20middle%20e a st/united%20kingdom/our%20insights/the%20arts%20in%20the%20uk%20seeing%20the %2 0big%20picture/the-arts-in-the-uk-seeing-the-big-picture_final.pdf

McLeod, S. (2023). Social identity theory in psychology. [online] Simply Psychology. Available at: https://www.simplypsychology.org/social-identity-theory.html.

Moreincommon (2025). This Place Matters: A social connection map of Britain. [online] Available at: https://www.moreincommon.org.uk/latest-insights/this-place-matters-a-social-connection-m a p-of-britain/.

Moss, K.L. (2010). Cultural Representation in Philadelphia Murals: Images of Resistance and Sites of Identity Negotiation. Western Journal of Communication, 74(4), pp.372–395. doi:https://doi.org/10.1080/10570314.2010.492819.

National Centre for Social Research. (2025). BSA 42 | Politics and social media | National Centre for Social Research. [online] Available at: https://natcen.ac.uk/publications/bsa-42-politics-and-social-media

Nezi, R., Karyotis, G. and Makropoulos, I. (2023). GreeSE Papers Hellenic Observatory Discussion Papers on Greece and Southeast Europe Culture wars? Assessing the impact of affective polarisation on cultural battles. [online] Available at: https://www.lse.ac.uk/Hellenic-Observatory/Assets/Documents/Publications/GreeSE-Paper s/ GreeSE-No190.pdf

Perkins, R., Mason-Bertrand, A., Tymoszuk, U., Spiro, N., Gee, K. and Williamon, A. (2021). Arts engagement supports social connectedness in adulthood: findings from the HEartS Survey. BMC Public Health, [online] 21(1). doi:https://doi.org/10.1186/s12889-021-11233-6.

Putnam, R. (2000). Bowling Alone: The Collapse and Revival of American Community. New York: Simon & Schuster.

Regehr, K., Shaughnessy, C., Zhao, M., Cambazoglu, I., Turner, A. and Shaughnessy, N. (2025). Normalizing toxicity: the role of recommender algorithms for young people's mental health and social wellbeing. Frontiers in Psychology, 16. doi:https://doi.org/10.3389/fpsyg.2025.1523649.

Rezaei, S., Childress, A., Kaul, B., Kelly Magill Rosales, Newell, A.D. and Rose, S. (2023). Using Visual Arts Education and Reflective Practice to Increase Empathy and Perspective Taking in Medical Students. MedEdPORTAL, [online] 19(11346). doi:https://doi.org/10.15766/mep_2374-8265.11346.

Schedler, A. (2023). Rethinking Political Polarization. Political Science Quarterly, [online] 138(3), pp.335–359. doi:https://doi.org/10.1093/psquar/qqad038.

Skinner, G. (2025). 85% of Britons continue to believe that Britain is divided as over half think differences in people's political views are so divisive that it is dangerous for society. [online] Ipsos. Available at: https://www.ipsos.com/en-uk/85-britons-continue-believe-britain-divided-over-half-think-differ ences-peoples-political-views-are

Srivastava, A. (2025). Investigating The Psychological Effects Of Algorithmic Echo Chambers On Belief Rigidity And Ingroup Bias Among Adolescents. IOSR Journal Of Humanities and Social Science (IOSR-JHSS, [online] 30(11). doi:https://doi.org/10.9790/0837-3011015257.

Stewart, I. (2024). 2024 general election: Performance of Reform and the Greens. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/2024-general-election-performance-of-reform-and-the greens/

Tajfel, H. and Turner, J.C. (1979). An integrative theory of intergroup conflict. [online] www.scirp.org. Available at: https://www.scirp.org/reference/ReferencesPapers?ReferenceID=757561.

Wright (2024) The UK in the World Values Survey. [online] The UK in the World Values Survey. Available at: https://www.uk-values.org/news-comment/polarisation-and-partisanship-10-key-takeaways-fr om-our-world-values-survey-conference.

Yilmaz, M. (2009). The social identity dimension of inter-group conflicts [online] Available at: https://www.ajindex.com/dosyalar/makale/acarindex-1423873606.pdf

Scrolling to the ballot box: assessing social media's impact on the UK's lowered voting age

Jozef Jorritsma

Abstract

Hoping to reenergise UK democracy, the Labour Government made a pledge to lower the voting age: for the first time in British history, suffrage will be extended to 16- and 17-year-olds. Their enfranchisement coincides with an era of digitalisation like no other before. Digital media has become integral to young people's socialisation, and is no less involved in their politicisation. As they form political identities, online platforms may threaten young people's engagement with politics. Conversely, social media may rejuvenate civic participation and enable young people to embrace their new suffrage as the policy intends. There is a need to evaluate both potential outcomes. This article finds that social media will not undermine the upcoming voter generation's voice. However, the policy's success hinges on governmental action to tackle misinformation, echo chambers, and harmful algorithms, at the risk of otherwise enabling online platforms alarming access to the latest generation of voting citizens.

Introduction to policy

Since the 2014 Scottish Independence Referendum, the minimum voting age across the UK has been split by devolved boundaries. Scotland was the first to lower voter age temporarily in 2014, but passed a resolution the following year to extend the vote to 16-year-olds across all Holyrood elections (NPR, 2014). The support for the extension was rooted in its measured success. Aiming to stimulate political participation, the 2014 referendum saw 21% more 16-17- year olds vote than the overall average turnout (Scotsman, 2024). Wales would follow in the Senedd and Elections Act 2020, demonstrating the widespread support for the reform. The Lib Dems have long campaigned on the issue (2001), and it was taken up by Miliband's Labour Party (BBC, 2013). Featured in their 2024 manifesto, the pledge for this electoral reform has finally reached Downing Street. Its primary aim is 'increasing the engagement of young people in our vibrant democracy' (Labour, 2024). At the time of this writing, the Representation of the People Act 2026 [RoPA] has passed its Second Reading. Its enactment will see the manifesto pledge, at last, materialise. Parliamentary discontentment, vocalised by Conservative and Reform opposition, highlights legitimate concern. Ironically, their argument revolves around the platforms they themselves increasingly integrate into campaign strategies, crediting to Reform's meteoric rise. Indeed, their concern is around social media. With 95% of 13-15- year olds in the UK using social media, its influence on the next voting population is undeniably reaching (Ofcom, 2025).

There is considerable existing scholarship on social media as a source of misinformation (Wu et al., 2019) and an 'echo-chamber' of assenting views (Barberá, 2020 ; Hale, 2024). Despite this, assessments rarely outright condemn social media without recognising its democratic benefits. Boulianne (2015) argues that it can stimulate political engagement,

strengthening democracy through increased civic participation. Despite the vast scholarly attention, social media evolves faster than academics and legislation. With the lowering of the voting age there is a renewed need for a review of social media's effects on young people in the political sphere.

This article aims to evaluate the lower voting age in relation to social media's influence, and answer the following questions. What barriers will social media present to the policy's success? Conversely, what benefits might social media generate? Will it threaten, or strengthen, the democratic legitimacy of the emerging voter block? And what governmental action is required to neutralise its political dangers for 16- and 17- year-olds, while heightening its rewards for UK democracy?

Policy Problem

In order to analyse any policy, we must start with the problem it aims to solve.

Hay (2007, p.12) observes that 'no single issue has prompted greater concern… in the broad literature on political disaffection than the question of voter turnout.' If we imagine UK democracy as a living organism, a stream of ballots would represent its lifeblood, and its heart would beat every election cycle. Western Democracies across the hemisphere are presented with a political participatory crisis. Declining formal political participation threatens the UK's democratic legitimacy and, if it continues to diminish, it 'will not prove a rapid death' (Hay, 2007, p.13).

Curtice and Seyd (2003) reported that Britain's crisis was evidenced by the latest election's formidably low turnout of 58%. Over two decades on, the 2024 election produced similarly concerning results. Turnout increased a mere 1.7% since Blair's 2001 victory and decreased by 7.6% since Johnson's 2019 success (BBC, 2024). The general decrease across all voter blocks demonstrates electoral participation to be an issue that supersedes voter demographics. Despite this national trend, UK political disinterest has been thought to be most prevalent amongst young people, especially those from disadvantaged backgrounds (Wunderlich, 2025).

Pressure group Unlock Democracy reported a 336% drop in 18 to 25-year-old's electoral enrolment applications during the regulated campaign period for 2025 local elections in comparison to 2024 (Parliament Politics Magazine, 2025). This issue is not limited to the

UK. The 2024 US presidential election saw 45% of 18-year-olds registered to vote, in contrast to over 75% of citizens aged 45 and over (Brill, 2025). Age-based suffrage inequality is likely to persist: Civic Centre's projections estimate 2.2 - 3.2 million unregistered 18- year olds in the 2026 midterms. In 2022, 60% of 18 to 19-year-olds were on the UK's electoral register, shadowed by 96% of those aged 65+ (Electoral Commission, 2026). Such data tempts the assumption that Civic's US projections might apply to UK politics. The participatory crisis of formal modes of politics, whilst not exclusive to young people, disproportionately disenfranchises the youth.

Statistical analysis, alongside a plenitude of supporting literature, acknowledges the attenuating trend of formal political participation. Yet scholars diverge on specific reasoning behind political disinterest for young people, and the effectiveness of policy solutions designed to solve (at least part of) the problem. The policy under review, a lower voting age, is one contested solution. The policy's goal, to stimulate young people's engagement in the formal political process of voting, arrives with no astonishment. Labour's policy method, to lower the voting age to 16, was more surprising. University of Exeter Professor, Dr Stuart Fox, led a study to measure the popularity of measures to engage young people in formal politics. The resulting observation showed Labour's manifesto promise to be strongly opposed by a third of subjects and strongly supported by only 10% (Maddern, 2024). One factor of this lack of support for lowering the voting age can be attributed to the influence of social media in young people's lives. Before honing in on this factor, we must first assess whether this policy alone may help solve the participation crisis, achieving its primary objective.

Looking ahead: how could lowering the voting age help solve the UK's participation problem?

Fervently advocating for the policy, Labour MP Vicky Foxcroft cited a Columbia University study during the Second Reading of the RoPA (2026). The study, summarised by Coppock and Green (2016), proved that 'the act of voting is itself habit forming.' Foxcroft reasons that a lowered voting age enables citizens to form habitual relationships to electoral participation earlier on (House of Commons, 2026). Accordingly, newly franchised 16- and 17- year-olds will have a higher propensity to persistently vote in elections than citizens who are accustomed to the voting process at later stages of life (Coppock & Green, 2016). This relies

on newly franchised young people voting in the first elections they are presented with. Some argue that the US study's transferability into UK politics is undermined by dominantly American electoral issues. Namely, voter fatigue from high frequency US elections can depress turnout (Garmann, 2016 ; Davenport et al., 2010). However, that is not to say the study's results provide no use for this article's evaluation.

Looking behind: how lowering the voting age has already worked out

We can turn to Scotland and Austria for empirical support of habit-forming voter theory among 16 to 17-year-olds. Since their 2014 electoral reform, Scotland has maintained higher electoral engagement among first-time voters enfranchised at 16 or 17 than those first granted the vote at 18 or older, evidenced by the 2021 Scottish Parliament elections (Eichhorn & Hübner, 2025). Such evidence supports habit-forming voter theory, though analysis awaits more election cycles to statistically prove its longevity in Scotland. Austria was the first European nation to extend the vote to 16-year-olds, in 2007. The subsequent 2008 national election was met with an 88% turnout among first time voters aged 16, successfully demonstrating young people's political keenness when presented with the ballot box. (Austrian Parliamentary Correspondence, 2018) The novelty effect did, however, wear off, declining below the national average at 63% during the 2013 election cycle (Schmidt & Edthofer, 2018). The Austrian Society for European Politics attributes civic education and outreach to the sustained voting tendencies of Austrian youth.

The UK Government should not overlook the combination of mechanisms involved in Austria's self-proclaimed success. Civic education is a crucial component of Austria's policy landscape, and, to replicate their results, widespread civic education, targeted at the emerging voter class, could be implemented across UK secondary schools. Attempts to promote active citizenship through educational means should be welcomed, albeit with caution. Incorporating civic education into national curricula presents challenges, such as promoting political participation without political biases. However, the UK Government cannot expect electoral reform alone to solve the participatory crisis. With monitored administration, cumulative civic education has already been shown to be positively associated with youth political participation (Weinberg, 2022). Social media's role in facilitating civic education may become a key determiner of the policy's success. Platforms provide an alternative to conventional classroom education, and, when used to promote active citizenship, has the potential to maximise the policy's aims to boost young people's electoral turnout.

Taking inspiration from Austria, such an objective is achievable in the UK. This article therefore assesses whether social media will contribute to its success. Alternatively, social media's harmful implications present barriers to the policy's goal which necessitate equal consideration. The purpose of this article is to take an evaluative role to deduce the best next steps to minimise social media's risks, whilst maximising its benefits, in relation to the prospective success of the policy under review.

How might social media help improve the policy's chance of success?

Improving democratic discourse

Arguably the most politically influential social media platform has been X, formerly known as Twitter. Musk's purchase in 2022 was motivated to secure, in his own words, 'a common digital town square' for 'the future of civilisation.' The so-called town square Musk speaks of can relate to Habermas' public sphere theory. It suggests that a third space, in this case, X, enables democratic discourse, fundamental to deliberative democracies (Goode, 2005; Jenkins, 2006). Social media largely provides this function: it enables users to participate directly with politics by sharing opinions and commenting on others. As a result, social media becomes a global 'Speakers Corner', enabling dialogue that is crucial to healthy democracies. These platforms can reenergise UK politics, boosting civic engagement, the purpose of lowering the voting age, by enabling political discourse. As such, social media can work in symphony with the policy to boost Britain's democracy and empower young people to civically participate in this 'town square'.

The reality of democratic discourse on social media

Scholars are not so idealistic. Although research recognises social media's positive capacity for dialogue, this is often overshadowed by the undemocratic mechanism for circulating discourse (Kruse et al., 2018). Selected content is algorithmically served to be compatible with individual political palettes, consequently forming echo-chambers across feeds. This fragments user bases, challenging categorisations of social media as a unified public sphere (Eisenegger & Schäfer, 2023). Milli et al. (2025) monitored Twitter's response to user engagement, finding that its algorithm amplified emotionally charged, hostile content that made users 'feel worse about their political out-group.' Such algorithms contribute

immensely to political polarisation, making controversial content suitable vehicles for dividing politics. Echo-chambers distort social media's democratic function, shrouding young people in divisive content that alters political perceptions.

Young people using platforms to enter politics earlier, due to a lowered voting age, are confronted by two risks. The first is that such interactions can deter further political engagement. Seeing hateful rhetoric amass virality more so than politically unifying content, politics may repulse young people, hindering the policy's aim to increase electoral participation. Secondly, young people's political preferences become obscured as adolescents are algorithmically led down popular rabbit holes, some of which may grow support for radicalism at the ballot box. Viral polarising content politically orientates emerging voters around popularity and divisiveness. Such trend-based politics often hijacks political issues and works in favour of populists. Political personality cults turn voters into followers, politicians into influencers, and election results into like-counts.

Trump's MAGA loyalists, enticed by provocative media, illustrate the consequences of digital populism. When political alliances are formed through virality rather than legitimate discourse, the objective to boost democracy is undermined. The method, enfranchising 16-17-year-olds, encourages influencer-politicians to target young people. However, given that over 50% of MAGA supporters are over 65- years old, there is scepticism whether politicians will target young voters any more than they already do with older citizens. The same can be said for the difference in effectiveness of such targeting. Although peer influence has long been considered a hallmark of young people's psychosociology (Brown, 2004), which may contribute to online radicalisation, the position that older voters are immune to similar traits has become laughable by the likes of the MAGA movement. On the contrary, older generations have been found to often be overlooked in confrontations of online extremism and prone to misinformation consumption online (Lang, 2025) This demonstrates social media's political harm across generations: online radicalisation has no age limit. Thus, it is crucial to not overlook older users once more.

At the very least, research suggests that all age groups are able to fall prey to social media harm. It is, therefore, unfair to direct concern exclusively at young people. Nevertheless, governments should continue to prioritise the upcoming generation of voters in social media policy. This is due to stronger paternal obligations for states to nurture young citizens, as well as nonpartisan public support for action targeted at the youth. Therefore, assessments of a

lower voting age should address that social media's harm is indiscriminate of age, whilst maintaining a prioritization of young people in policy recommendations. As already stated, older users are prone to consume online misinformation (Lang, 2025), though the same can be said for young people. Since provocation is a vehicle for virality, algorithms tend to distribute misinformation more widely due to their, often emotionally charged, novelty. (Vosoughi et al., 2018).

(Mis)information on social media

The government aims to help solve the crisis amongst young people's political participation by lowering the voting age. In digital spaces, where false political information thrives, the crisis grows as participation is based on unrealities. Political opinions rooted in counter-factual claims surface, and voters' true intentions are obscured. Engagement-based algorithms disseminate unverified information through echo-chambers, whilst TV news has been found to expose audiences to a more robust range of political perspectives (Prior, 2007). The effect is concentrated amongst young people: 8% of those 55 or above use social media to source political news, in contrast to 58% of 35- year olds and lower (Kurtice et al., 2025).

Since adolescents have highly malleable social identities (Karantzis, 2018), social media becomes a force that impresses upon 16-17-year-olds' ideological alignments. Entering a critical phase of political socialisation (Kops & Wachs, 2025), it provides access for unreliable externalities, such as misinformation, to shape young people's political identities. Despite previous opposing findings, research suggests that young people often overestimate their ability to recognise misinformation online. In a 2022 report, Ofcom documented that only 11% of 11-17- year olds could reliably identify the signs that a post was genuine, whilst a separate study found that over 90% of social media users encountered misinformation online (Alan Turing Institute, 2024). Political misinformation seeks to base real politics on fugazi narratives, whilst affecting young people as the primary consumers of political information on social media. Thus, social media poses a threat that undermines the policy's goal to boost UK democracy through amplified political misinformation. By gaining suffrage, 16- 17- year olds may be targeted as impressionable victims of political misinformation equipped with a useful voting right. Social media, therefore, presents a risk of intoxicating politics with false information through a lowered voting age.

A recent rising concern surrounding misinformation lies in generative AI. As it enhances in technicality and accessibility, counterfeit content is often indistinguishable from reality. Branding AI posts as real-life contributes to the erosion of trust in public media. As a result, young people may become disillusioned with social media content. This makes it less likely for users to utilise the ability to platform and engage with accurate information online. It is difficult to ignore endless accusations of AI that often flood comment sections under unverified posts. This creates a 'boy who cried wolf' effect, whereby such false claims of AI diminish truthful accusations. Feeds clouded by fiction also disillusion users from such truths. As an extension, young people may become digitally disengaged with politics, unknowing of what can be used to factually base individual political identities. Consequently, AI holds the capacity to threaten the emerging voting generation and reverse the objectives of lowering the voting age.

Nonetheless, we should not disregard the importance of information circulation altogether. The distribution of information by individuals on social media has the unique capacity to connect the globe, and that is a democratic benefit hard to overstate. Although misinformation can achieve virality, so can real-time news uploaded incessantly; fact-checked sources; first-hand testimonies. Social media provides a unique speakerphone that is used to voice a multitude of political victims and actors. On the same day their victims documented their narrowing escape from Kunduz, Taliban selfies began circulating on various platforms (BBC, 2020), profoundly exemplifying the political range of information posted online. This frequently results in a digital mobilisation of support, or dissent, that can transmogrify into social movements with political consequences. There is no need to look further back than the social upheaval that erupted from the killing of George by US police in 2020 (BBC, 2020). Twitter experienced 47.8 million uses of #BlackLivesMatter within two weeks of a viral upload of Floyd's death (Hughes, ISRM). The movement was characterised by a surge of civic engagement led dominantly by young people. Being plugged in to social media therefore enables access to raw news that drives political engagement, helping achieve the aims of the policy under review.

Theorising the reasons behind low voter turnout amongst young people, Pattie et al. (2003) argues that 'the less well-resourced are unlikely to participate in civic life.' This contrasts with increased voter turnout of educated citizens (Hay, 2007). The resulting effect is 'the voice of the less privileged… becoming excluded from the participative, consumerist,

atomised politics of the twenty-first century' (Pattie et al, 2003) Social media counteracts such phenomena: it provides educational resources across all voter blocks. This includes news reporting and awareness of political events, allowing widespread communicative access for voter mobilisation. Accessible news reporting and widespread awareness of political events allows for networks of communicative access for voter mobilisation. Through circulating vital information and providing a forum for discourse, social media may work in partnership with a lowered voting age to encourage formal political participation, as well as enable informal participation through the platform itself.

A lower voting age for the social media generation?

The 2026 Representation of the People Act is due for its committee stage this April: it is increasingly certain that 16- year olds will be able to vote in the next general election. This article has shed light on the consequences social media poses in relation to this policy, but what can the government do to minimise such risks and increase its benefits?

Policy aimed to tackle misinformation and reduce harmful discourse is like an acrobat on a seesaw. In order for social media to function as a public sphere, regulation must be balanced in its approach. It must verify and remove false information without encroaching on the user's right to unconstrained speech. All the while, governments don't seek to make enemies with media giants and the so-called tech bros, putting governments in a tricky position. Either way, regulation is bound to feel as though it is flipping someone off. Despite this, governments should not be dissuaded into doing nothing; the lowered voting age should signal new provisions for such regulations. With delicate complications, Australia's blanket ban for under 16s presents a tempting solution to social media's problems. However, this nuclear option presents a conflicting message to those about to enter the voting age. Enfranchisement aims to extend the political involvement of young people, working at odds with restrictions of vital forums of online discourse as well as accessible information through a blanket ban. Australian youth responded to their government's policy by bypassing blocks through VPNs and other methods, including scrunching up faces to appear older for age confirmation (Jaen, 2026). Overall, an all-out ban seems counterproductive.

Perhaps a better solution lies in current legislation. A parliamentary committee observed that the Online Safety Act 2023 (OSA) which, among other things, granted Ofcom fining power over misinforming platforms, was not being implemented to its fullest (Science, Innovation and Technology Committee, [SIT] 2025). Whilst Governments should be wary when

approaching this policy seesaw, independent companies, like Ofcom, can ease the pressure. As an independent regulator, it is crucial that it feels public pressure to thoroughly fine platforms caught circulating misinformation amongst young people. The same report also found that the OSA was 'not designed to tackle misinformation,' calling for fresh legislation on the matter (SIT, 2025). Legislation has not kept up with evolving media. For instance, Starmer's government should respond to the EU Commission's report of TikTok's AI-enhanced algorithms and infinite scroll features. With little requirements for transparency over harvesting users' data, as well as its mechanical design to be 'addictive' (EU, 2026), there is no robust legislation that prevents echo-chambers from being engrained into young people's feeds. Tackling this issue would allow for sensible regulation of misinforming soundbite politics. Restricting social media's addictive tools would reduce echo-chambers that can lead to radicalism and social fragmentation. Such an aim to stimulate social media's benefits would work complimentarily to a lower voter age, minimising the harm social media can present.

Social media's benefits include providing a forum for democratic discourse, delivering important information to voters, and enabling communicative mobilisation. Removing misinformation and reducing echo-chambers might foster a culture of informative social media and healthy discussion that boosts deliberative democracy. Such enforcement also incentivises platforms themselves to promote accurate content to avoid fines or distrust. Ultimately, however, the policy is still very much on a seesaw. Media giants will not respond well to full scale intervention and there is an evident challenge to rely on Ofcom to implement regulation.

Looking beyond social media

Alternatively, policy could move away from social media altogether. Its positive qualities, like engineering a space for public discourse, present an underbelly of harm to its democratic functions through skewed algorithms and misinformation. The public sphere that social media fails to reliably fulfil, however, can be created without Wi-Fi signals or phone screens. Research suggests that neutral third spaces strengthen 'the social fabric of communities' by enabling participation of civic life (Thornton, 2024). There is no online requirement; social media does not have to be young people's third space. Instead, through investment into accessible third spaces like youth community centres, policy can support 16- year olds transitioning into voting citizens by fostering their political-socialisation and forming a sense

of social capital (Littman, 2021). This approach to a lowered voting age would almost certainly garner parental support, whilst creating a space for young people to form political identities free from social media's contamination.

Despite this, it is difficult to imagine that physical third spaces would replace social media use altogether. Since its harm would likely continue, other options to combat risks associated with social media use must be examined. Paterson (2026) claims that policy should focus on the demand dimension of social media content as well as its supply side. Targeted tools, such as 'pre-bunking', have been shown to effectively reduce risk of misinformation (Paterson, 2026) through psychological inoculation (Roozenbeek et al., 2022). In practice, this includes simple 30-second clips that pre-expose users to common manipulation tactics invoked through social media. In policy, this would contribute to safer uses of social media, thus benefitting a lowered voting age by reducing online risks. Such a policy should encourage young people to take a critical approach to digital media and enable skills that reduce platforms' political detriment to be built by young people early on. This would not dispel social media's harms entirely. However, when in operation with robust regulation and an increase of physical third spaces, policy can transform social media from a barrier to a healthier democracy, to the antidote to the UK's youth participation crisis.

Conclusion

To conclude, it is this article's view that social media will not prove detrimental to UK democracy after a lowered voting age, or at least no more than it already has done. Its problems are just as prevalent in the adult population than it is amongst youth. Moreover, older voters have been found to be more susceptible to fall for misinformation due to congeniality bias (Lyons, 2024). Since young people are yet to form firm political identities, older users are also at great risk of falling into echo-chambers of their pre-existing views. Social media's harm is not exclusive to young people, this has already been established. Nevertheless, a lowered voting age should ignite discussion of digital media's presence within young people's political lives. Such dialogue should motivate policymakers to address social media's challenges to young people in an overall effort to solve the youth participatory crisis. Finally granted a say in the election process, there is no better time for governments to work to minimise social media's harms to young people's politics, and empower their newfound voices as they transition into the formal political sphere.

Bibliography

The Alan Turing Institute (2024) More than 90% of the UK population have encountered misinformation online. New Research, May 30th. Available at https://www.turing.ac.uk/news/more-90-uk-population-have-encountered misinformation-online .

Barberá, P. (2020) 'Social Media, Echo Chambers, and Political Polarization', in Social Media and Democracy (eds. N. Persily & J.A. Tucker). Cambridge University Press, chap. 3, pp. 34–55. Available at https://www.cambridge.org/core/services/aop-cambridgecore/content/view/333A5B4DE1B67EFF7876261118CCFE19/9781108835558c3_34-55.pdf/social-media-echo-chambers-and-political-polarization.pdf .

BBC (2013) Labour conference: Ed Miliband calls for votes at 16. Sep 24th. Available at https://www.bbc.co.uk/news/uk-politics-24229366 .

BBC (2015) Taliban selfies: Why militants posed for photos in Kunduz. Sep 30th. Available at https://www.bbc.co.uk/news/world-asia-34387854 .

BBC (2020) Black Lives Matter: From social media post to global movement. July 10th. Available at https://www.bbc.co.uk/news/world-us-canada-53273381 .

BBC (2024) UK General election 2024 Results. Available at https://www.bbc.co.uk/news/election/2024/uk/results

Briatte, F., Kelbel, C., Navarro, J. (2024) Do (too many) elections depress participation? How the position, frequency and nature of domestic ballots affect turnout in European Parliament elections. European Union Politics. Vol. 25, No. 3, pp 549-568. Available at https://doi.org/10.1177/14651165241257779 .

Brill, L. (2025) These charts will change the way you think about youth voter turnout. The Civics Center, June 5th. Available at https://www.thecivicscenter.org/blog/youth-voter-turnout-charts .

Brown, B. (2004) ‘Adolescent’s Relationships with Peers’ in Handbook of Adolescent Psychology, Second Edition (eds. Lerner, R. & Steinberg, L), chap. 12. Available at https://doi.org/10.1002/9780471726746.ch12 .

Boles, S. (2026) Why are older adults more likely to share misinformation online? The Harvard Gazette, Jan 5th. Available at https://news.harvard.edu/gazette/story/2026/01/why-are-older-adults-more-likely to-share-misinformation-online/ .

Boulianne, S. (2015) Social media use and participation: a meta-analysis of current research. Information, Communication & Society. Vol. 18, No. 5, pp 524- 538. Available at https://www.tandfonline.com/doi/full/10.1080/1369118X.2015.1008542 .

Coppock, A. & Green, D. (2016) Is Voting Habit Forming? New Evidence from Experiments and Regression Discontinuities. American Journal of Political Science, Vol. 60, No.4. pp 1004-1062. Available at https://www.jstor.org/stable/24877471 .

Curtice, J. and Seyd, B. (2003) Is there a crisis of political participation? Chapter appears in: Park, A. et al, British Social Attitudes: the 20th Report – Continuity and Change over Two Decades, London. Sage. pp 93-107. Available at https://kar.kent.ac.uk/958/1/BSA2003.pdf .

Curtice, J., Mezzanzanica, M., & Bailey, K. (2025) Politics and social media: Scrolling politics - engagement, trust and polarisation in the digital age. National Centre for Social Research: Politics and social media. BSA 42. Available at https://natcen.ac.uk/publications/bsa-42-politics-and-social-media .

Davenport, C., Gerber, A., Green, D., Larimer, C., Mann, C. and Panagopoulos, C. (2010) The Enduring Effects of Social Pressure: Tracking Campaign Experiments Over a Series of Elections. Political Behaviour. Vol. 32, No. 3, pp 423– 30. Available at https://www.jstor.org/stable/40960946 .

Eichhorn, J. & Hüebner, C. (2024) Longer‐Term Effects of Voting at Age 16: Higher Turnout Among Young People in Scotland. Politics and Governance. Vol. 23, No. 9283. Available at https://www.cogitatiopress.com/politicsandgovernance/article/viewFile/9283/4184 .

Eichhorn, J. & Hüebner, C. (2020) Lowering the Voting Age to 16. Palgrave Studies in Young People and Politics. Chap. 7, pp 121-142. Available at https://www.sps.ed.ac.uk/sites/default/files/assets/doc/Votes%20at%2016%20in%

20Scotland.pdf .

Eisenegger, M. & Schäfer, M. (2023) Reconceptualizing public sphere(s) in the digital age? On the role and future of public sphere theory. Communication Theory. Vol. 33, No. 2-3, pp 61-69. Available at https://doi.org/10.1093/ct/qtad011 .

Elections 2025: Less than half as many young people have registered to vote for May elections (2025). The Parliament Politics, April 11th. Available at https://parliamentnews.co.uk/elections-2025-less-than-half-as-many-young-people-have-registered-to-vote-for-may elections.

Electoral Commission (2026) Explore the data: Who is and isn't registered to vote? Electoral registration research. Available at https://www.electoralcommission.org.uk/research-reports-and-data/electoral registration-research/explore-data-who-and-isnt-registered-vote .

European Commission (2026) Commission preliminarily finds TikTok's addictive design in breach of the Digital Services Act. Press release, Feb 6th. Available at https://ec.europa.eu/commission/presscorner/detail/en/ip_26_312 .

European Commission (2025) Youth participation in representative democracy. Chap. 5, 2. Available at https://national policies.eacea.ec.europa.eu/youthwiki/chapters/austria/52-youth-participation-in representative-democracy .

Fell, R. (2024) What impact did lowering the voting age have for Scottish elections? What will Keir Starmer's election move mean? Scotsman Article, June 2nd. Available at https://www.scotsman.com/news/politics/what-impact-did lowering-the-voting-age-have-for-scottish-elections-what-will-keir-starmers election-move-mean-4649171 .

Garmann, S. (2017) Election frequency, choice fatigue, and voter turnout. European Journal of Political Economy. Vol. 47, pp 19-35. Available at https://api.semanticscholar.org/CorpusID:157217694 .

Goode, L. (2005) Jurgen Habermas : Democracy and the Public Sphere. Pluto Press. Available at https://ebookcentral.proquest.com/lib/york ebooks/reader.action?docID=3386163&c=UERG&ppg=1 .

Hale, M. (2024) Amplifying Discord or Fostering Freedom? Contrasting Implications of Social Media Echo Chambers for Deliberative and Libertarian Conceptions of Democratic Free Speech. Adelaide Law Review. Vol. 45, No. 3, pp 527-553. Available at

https://heinonline.org/HOL/Page?collection=journals&handle=hein.journals/adelrev45&id=547&men_tab=srchresults .

Hay, C. (2007) Why we hate politics. Cambridge Polity.

House of Commons. (2026) Representation of the People Bill: Second Reading. Hansard, Vol. 781, March 2nd. Available at https://hansard.parliament.uk/commons/2026-03-02/debates/702183A9-7B57-487C-9634-BF459DB65DB2/RepresentationOfThePeopleBill .

Hughes, S. (2022) Social Media Case Study: The Killing of George Floyd. ISRM blog. Available at https://www.theisrm.org/social-media-case-study-the killing-of-george-floyd/ .

Karantzis, C. (2018) A Most Impressionable Age. Socioemotioanl and Psychological Development in Adolescence blog, Oct 6th. Available at https://wp.nyu.edu/urbanyouthnyu/2018/10/06/a-most-impressionable-age/ .

Khaled, F. (2022) Why Did Elon Musk Buy Twitter? Newsweek. Oct 27th. Available at https://www.newsweek.com/elon-musk-reveals-why-bought-twitter 1755180 .

Kops, M., Schittenhelm, C. & Wachs, S. (2025) Young people and false information: A scoping review of responses, influential factors, consequences, and prevention programs.

Computers in Human Behaviour. Vol. 169, No. 1008650. Available at https://doi.org/10.1016/j.chb.2025.108650 .

Kruse, L., Norris, D. & Flinchum, J. (2018) Social Media as a Public Sphere? Politics on Social Media. The Sociological Quarterly. Vol. 59, No.1, pp 62-84. Available at https://www.jstor.org/stable/26802985 .

Labour (2024) Change – Manifesto. Available at https://labour.org.uk/wp content/uploads/2024/06/Labour-Party-manifesto-2024.pdf

Lang, I. (2025) When Grandpa Turns Extremist: Digital Immigrants, Radicalisation and the German Reich Citizens Movement. Global Network on Extremism and Technology. Available at https://gnet-research.org/2025/03/31/when-grandpa-turns-extremist-digital-immigrants-radicalisation-and-the-german-reich-citizens-movement/ .

Liberal Democrats (2001) Freedom, Justice, Honesty – Manifesto. Available at https://general-election-2010.co.uk/2001-general-election-manifestos/Liberal

Democrats-Party-Manifesto-2001.pdf .

Littman, D. (2021)How Can “Third Place” Settings Support Young People who Endure Social Marginalization? American Journal of Community Psychology. Available at https://www.communitypsychology.com/third-places/

Lyons, B., Montgomery, J., Reifler, J. (2024) Partisanship and Older Americans’ Engagement with Dubious Political News. Public Opinion Quarterly. Vol. 88, No. 3, pp 962-1097. Available at https://doi.org/10.1093/poq/nfae052 .

Maddern, K. (2024) Votes at 16 and national service policies both fail to impress voters, new study shows. University of Exeter News, June 22nd. Available at https://news.exeter.ac.uk/faculty-of-humanities-arts-and-social-sciences/votes-at 16-and-national-service-policies-both-fail-to-impress-voters-new-study-shows/ .

Milli, S., Carroll, M., Wang, Y., Pandey, S., Zhao, S. & Dragan, A. (2025) Engagement, user satisfaction, and the amplification of divisive content on social media. PNAS Nexus. Vol. 4, No. 3. Available at
https://doi.org/10.1093/pnasnexus/pgaf062 .

Musk, E. (2022) X. Oct 27th. Available at
https://x.com/elonmusk/status/1585619322239561728 .

Ofcom (2022) ‘Understanding the Online Space’ in Children and parents: media use and attitudes report, pp 40-50. Available at
https://www.ofcom.org.uk/siteassets/resources/documents/research-and data/media-literacy-research/children/childrens-media-use-and-attitudes 2022/childrens-media-use-and-attitudes-report-2022.pdf?v=327686 .

Ofcom (2025). Children and Parents: Media Use and Attitudes Report. Available at
https://www.ofcom.org.uk/siteassets/resources/documents/research and-data/media-literacy-research/children/childrens-media-use-and attitudes-report-2025/childrens-media-literacy-report

Paterson, J. (2026) Australia’s social media ban is only half the solution for online radicalisation. The interpreter, Feb 12th. Available at https://www.lowyinstitute.org/the-interpreter/australia-s-social-media-ban-only-half-solution-online-radicalisation.

Pattie, C., Seyd, P. & Whiteley, P. (2003) Citizenship and Civic Engagement: Attitudes and Behaviour in Britain. Political Studies. Vol. 51, No. 10, p 633. Available at https://doi.org/10.1111/1467-9248.00435.

Parliament of the Republic of Austria (2018) Wählen mit 16 – ErstwählerInnen bei NR-Wahl 2017. Parliamentary Correspondence [Parlamentskorrespondenz]. No.145, February 23rd. Available at https://www.parlament.gv.at/aktuelles/pk/jahr_2018/pk0145/.

Prior, M. (2007) 'Conditional Political Learning', in Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections. Cambridge University Press. Chap. 2, pp. 27–52. Available at https://www.cambridge.org/core/books/postbroadcast-democracy/conditional political-learning/32D8065D8705FDF0CBD3F55802DAA82B.

Roozenbeek, J., Van Der Linden, S., Goldberg, B., Rathje, S. & Lewandowsky, S. (2022) Psychological inoculation improves resilience against misinformation on social media. Science Advances, Vol 8, No 34. Available at 10.1126/sciadv.abo6254.

Schmidt, P. & Edthofer, J. (2018) Voting at 16 in Austria: a possible model for the EU? The Austrian Society for European Politics [ÖGfE]. July 12th. Available at https://www.oegfe.at/en/statements/12_elcano_voting16/.

Science, Innovation & Technology Committee (2025) Social media, misinformation and harmful algorithms: Government and Ofcom responses. London: House of Commons, 137. Available at https://publications.parliament.uk/pa/cm5901/cmselect/cmsctech/441/report.html.

Shapiro, A. (2014) For Scotland's 16-Year-Olds, The First Vote Will Be On Independence. NPR. Politics & Policy, Sep 17th. Available at https://www.npr.org/sections/parallels/2014/09/17/349085178/for-scotlands-16-year-olds-the-first-vote-will-be-on-independence.

Thornton, B. (2024) The Emergence of Digital Third Place & the Waning Role of Physical Third Place. Pittsburg State University, Masters. Available at file:///C:/Users/jozef/Downloads/BlayneThorntonMastersThirdPlacePaper.pdf.

Vosoughi, S. & Aral, S. (2018) The spread of true and false news online. Science. Vol. 359, No. 6380, pp 1146-1151. Available at https://doi.org/10.1126/science.aap9559.

Wu, L., Morstatter, F., Carley, K., Liu, H. (2019) Misinformation in Social Media: Definition, Manipulation, and Detection. ACM. SIGKDD Explorations Newsletter , Vol. 21, No. 2, pp 80-90. Available at https://doi.org/10.1145/3373464.3373475.

Wunderlich, L. (2025) Disinterested and Disillusioned? Information and Political

Engagement Practices of Young People From Disadvantaged Backgrounds. Sage: Youth & Society. Vol. 57, No. 7, Feb 7th. Available at https://journals.sagepub.com/doi/10.1177/0044118X251315206.

North Korea: The Paradox of a (Dis)Connected Society

Rafael Colman-Chadwick

Abstract

Frequently portrayed as the world's most isolated state, North Korea — officially the Democratic People's Republic of Korea (DPRK) — is diplomatically estranged, economically sanctioned, and informationally sealed. However, such descriptions only partially capture the nature of its social and political order. This report therefore conceptualises North Korea as a (dis)connected society, defined by a fundamental paradox: while the state remains externally disconnected from the global system, it is internally hyper-connected through dense networks of ideology, continuous surveillance, and state mobilisation.

Defining Disconnection

Disconnection, in this context, does not simply denote absence from international institutions; rather, it refers to a deliberate structuring of society that limits external flows of information, capital, and political influence while intensifying vertical bonds between citizen and regime. For the purposes of this paper, two pillars of (dis)connection are examined in relation to North Korea: isolation and economic insecurity, where the state selectively focuses on weaponry, rather than increasing living standards on an already burdened population. These dynamics are not incidental failures of governance but structural features that sustain regime stability. At the same time, they generate profound human insecurity and limit prospects for reintegration. By analysing these tensions, the report seeks to assess whether carefully managed forms of reconnection — including humanitarian engagement and diaspora integration — could mitigate alienation without destabilising regional security.

Isolation

Internal Closure and External Coercive Reach

Although North Korea's isolation is often presented as a geographic and diplomatic anomaly, this framing obscures its purposeful political design — a socio-political architecture intended to restrict external influences while simultaneously intensifying internal control. North Korea has embedded these dynamics by enforcing strict limits on information access, movement, and political expression, thereby creating the foundations of a dense surveillance apparatus that shapes daily life.

Since the COVID-19 pandemic, eight escapees from North Korea described to the Transitional Justice Working Group (TJWG) and Human Rights Watch that there were

“restrictions on movement; and limited access to food, medicine, and products like fertiliser, soap, or batteries after the government sealed its borders in 2020” (Human Rights Watch, 2025). These escapees further reported that “the government intensified its ideological control and surveillance, targeting people for seeking out foreign media, and failed to adopt measures to mitigate the economic impact of its restrictions” (Human Rights Watch, 2025). According to Kim Il-hyeok, a rice trader from South Hwanghae province who defected in May 2023, he had reportedly seen no executions between 2012 and 2019. However, he reported witnessing “people shot every two months in 2020” (Human Rights Watch, 2025). He further stated that between 2022 and 2023 there were three executions by firing squad every two months. This cocktail of mechanisms illustrates the regime’s systematic enclosure of society, where isolation functions as a structural pillar of authoritarian stability. Importantly, the eight escapees referenced in these testimonies originated from multiple regions of North Korea, indicating the breadth of its control across the state.

This internal closure does not imply that the regime withdraws from the external world. Like many authoritarian governments, North Korea selectively engages with transnational spaces when it perceives threats to its own stability. Authoritarian states increasingly project coercive power beyond their borders to monitor, intimidate, and deter dissent or defection — a phenomenon widely described as transnational repression (Michaelsen and Thumfart, 2023). In the case of North Korea, diaspora communities and defectors living abroad have been subject to a range of extraterritorial pressures that reveal the paradoxical nature of its isolation: the state isolates its population internally from information and influence, yet actively polices perceived political threats in global contexts.

Coercive Reach Beyond Borders

North Korea’s application of coercive tactics beyond its territorial boundaries demonstrates that isolation, for the regime, is not synonymous with withdrawal. When North Korea reaches beyond its borders to intimidate, silence, or coerce members of its diaspora and defector communities, it employs both direct and indirect strategies to suppress dissent abroad. The regime has historically sought to maintain control over those who flee, using intimidation, surveillance, and coercive pressure to constrain political mobilisation and silence critical voices.

Empirical research on North Korean defectors highlights the coercive environment surrounding those who attempt to leave the country. Survey evidence collected from over 1,300 North Korean refugees in China and 300 refugees in South Korea shows that approximately one-quarter of respondents reported being arrested in China and forcibly repatriated to North Korea, while nearly 10% reported incarceration in correctional or political detention facilities following repatriation (Haggard and Noland, 2009). Among those who had been detained, 90% reported witnessing forced starvation, 60% reported deaths caused by beatings or torture, and 27% reported witnessing executions, illustrating the severe, and sometimes fatal, consequences faced by defectors (Haggard and Noland, 2009).

While systematic studies specifically focused on North Korea's extraterritorial repression remain limited due to data constraints, broader research on authoritarian transnational repression demonstrates similar patterns whereby regimes use surveillance technologies, digital harassment, and coercion-by-proxy — such as threatening family members remaining in the home country — to influence behaviour abroad (Council of Europe, 2024; Tsourapas, 2021).

Human rights reporting further indicates that North Korea's efforts to influence defectors do not end once individuals leave the country. Government assessments note that North Korean authorities continue to monitor and intimidate defectors abroad, including through threats directed at family members who remain inside the country (U.S. Department of State, 2024). In many documented cases, relatives of defectors have faced interrogation, surveillance, or punitive measures as a result of contact with family members overseas, reinforcing the regime's capacity to exert pressure across borders (Freedom House, 2021). Such practices create a persistent climate of fear within diaspora communities, discouraging public criticism of the regime and limiting the willingness of defectors to participate in activism or advocacy.
A highly visible demonstration of this external reach occurred in 2017 with the assassination of Kim Jong-nam, the estranged half-brother of North Korean leader Kim Jong Un, at Kuala Lumpur International Airport. Kim Jong-nam was killed using the nerve agent VX in an operation widely attributed by investigators and international observers to agents acting on behalf of the North Korean state (Reuters, 2017). Beyond its immediate political implications, the killing illustrates the extent to which North Korea's mechanisms of control can extend into international space. The episode signalled that perceived political threats may

be targeted even outside North Korea's borders and served as a broader message to elites, defectors, and critics regarding the potential consequences of dissent or political rivalry.

Taken together, these dynamics suggest that North Korea's isolation infrastructure should not be understood as simple disengagement from the outside world. Rather, the regime maintains a selectively permeable strategy of control: domestically it constructs a highly regulated informational environment, while internationally it employs targeted forms of coercion intended to intimidate defectors, discourage diaspora activism, and protect regime stability. Through surveillance, threats to family members, and occasional high-profile acts of violence abroad, North Korea demonstrates that its political authority can extend well beyond its territorial boundaries.

Economic Insecurity

External Threats, Strategic Coercion and the Regime's Selective Connectivity

Through coercive statecraft, North Korea is frequently characterised as economically isolated; North Korea's economic strategy reflects a more complex pattern of selective connectivity, whereby the regime remains deeply integrated into global security dynamics through military development. Thus economic insecurity represents a central structural feature of North Korea's political economy.

For example, there is a sustained allocation of resources to strategic weapons programmes despite economic deprivation among the population. This dynamic illustrates how the regime prioritises coercive capabilities as they perceive them to be a key instrument of survival, due to the fact that they offer deterrence and international leverage. This idea is further reinforced by evidence that the country does not signal simple economic isolation. Consequently, North Korea's economic insecurity cannot be understood solely as the outcome of sanctions or underdevelopment; it is also shaped by deliberate policy choices that channel scarce resources into military programmes designed to counter perceived external threats.

Strategic Weapons Development and Military Prioritisation

Despite widespread economic constraints, North Korea has consistently prioritised the development of weapons systems that enhance its deterrence posture. International sanctions have been initiated to target North Korea's nuclear and missile programmes, significantly restricting the country's access to global markets and financial systems, thus contributing to economic stagnation and chronic shortages of essential goods. Evidence suggests that revenues generated through illicit networks, overseas labour, and state-controlled enterprises continue to be redirected towards weapons development (U.S. Department of the Treasury, 2024). These financial channels offer the regime sustenance in enabling costly military programmes, despite the fact that much of the population faces economic hardship.

Recent analysis also highlights the ongoing expansion of North Korea's nuclear capabilities, including the development of new warheads and fissile material production infrastructure. Studies of the Yongbyon nuclear complex suggest that the country's experimental light-water reactor could contribute to the production of tritium and plutonium for nuclear weapons, potentially supporting the development of multiple new warheads annually (Park and Glaser, 2024). Thus, such technological investments illustrate how strategic weapons programmes remain central to the regime's security doctrine. At the same time, they further reinforce economic insecurity domestically by diverting scarce resources away from civilian production and infrastructure.

North Korea has publicly revealed aspects of its uranium enrichment capabilities, signalling ongoing development within its nuclear programme. State media reported that Kim Jong Un toured two uranium enrichment facilities in September 2024 and January 2025, providing rare visual insight into the country's nuclear infrastructure. In September 2024, official images showed Kim inspecting an unspecified enrichment site, accompanied by references to an active construction area believed to be part of an expansion aimed at increasing production capacity. The reporting further emphasised his intention to expand the number of centrifuges, indicating a continued commitment to scaling fissile material production. Taken together, these disclosures suggest not only technological progression but also a strategic effort to demonstrate capability, reinforcing both domestic legitimacy and external deterrence.

This pattern reflects a broader logic of regime survival. Military capabilities provide North Korea with bargaining power in international negotiations, deterrence against perceived external threats, and leverage in geopolitical competition. Weapons development functions

not only as a security policy but also as an economic strategy through which the regime offsets structural vulnerabilities created by sanctions and diplomatic isolation (Preble and Willis, 2024).

Food insecurity, Regime Prioritisation and Structural Disconnection

One of the key structural features of the regime is chronic food shortages. International assessments indicate that a significant proportion remains undernourished, with persistent themes of malnutrition reflecting long-term deficiencies in food production, distribution and access. Yet, this is closely linked to the regime's deliberate prioritisation of strategic capabilities over civilian welfare. This is due to the cultural fact that this is a system reflecting regime survival, that is anchored in the belief that deterrence and coercive capacity are more valuable than economic performance (by lifting people out of starvation and poor living standards) and popular legitimacy. Thus, in this context, military strength is a structural necessity: as the regime's security strategy depends on credible threats that deter external intervention and provide leverage in international negotiations. This can be seen in the cases such as the Hanoi Summit in 2019 where Kim Jong Un met with Donald Trump to negotiate over the denuclearisation of North Korea.

Disconnection plays a central role in enabling this dynamic. As North Korea intentionally limits integration with global markets, restricting humanitarian access, and maintaining tight control over domestic distribution networks, the regime reduces external scrutiny and preserves its autonomy over reduced resource allocation. This suggests that food insecurity in North Korea is not merely an unintended outcome of isolation, but a structure that is structurally embedded within the regime's model of governance.

Biological Weapons as Strategic Asymmetry

Due to North Korea constantly aiming to expand its weapons arsenal, analysts have increasingly drawn attention to potential biological weapons capabilities. Just like the nuclear and missile programme, information remains extremely limited due to the country’s secrecy, yet recent assessments suggest that North Korea possesses the technical capability to produce a multitude of biological agents that could be weaponised for military purposes. According to research on bioterrorism preparedness, North Korea may be capable of developing biological

agents such as anthrax, plague, smallpox, and botulinum toxin, which could be delivered via missiles, artillery shells, or covert operations (Hwang, 2024). These capabilities represent a form of asymmetric warfare that allows the regime to compensate for conventional military and economic disadvantages.

From an economic perspective, biological weapons present a markedly lower barrier to entry than nuclear or advanced missile capabilities, making them an attractive asymmetric option for resource-constrained states such as North Korea. Academic and policy research consistently highlights that biological weapons are significantly cheaper to develop than nuclear weapons programmes, largely because they rely on dual-use infrastructure and widely accessible scientific knowledge rather than rare fissile materials and complex delivery systems (Cenciarelli et al., 2013; Stimson Center, 2023). In the context of economic insecurity, this relative affordability strengthens the strategic logic for pursuing biological capabilities as a means of offsetting financial and technological constraints while maintaining coercive potential. Moreover, biological weapons align with North Korea's broader structural disconnections from the international system, as their development can be concealed within domestic, dual-use scientific sectors, reducing reliance on external networks and reinforcing the regime's preference for strategic self-reliance and isolation. Further assessments by international monitoring bodies indicate that North Korea maintains what has been described as a "national-level offensive biological weapons programme" (U.S. Department of State, 2025).

Selective Connectivity and the Political Economy of Deterrence

The persistence of weapons development programmes illustrates how North Korea's economic insecurity is embedded within a broader strategy of selective connectivity. While the regime remains actively engaged in international security competition, it simultaneously restricts many forms of economic interaction with the outside world. Weapons development, sanctions evasion, and illicit financial networks enable the state to maintain a degree of external connectivity even as domestic economic conditions remain fragile. This dynamic is further reinforced by deepening strategic relationships with Russia, particularly following the 2024 Comprehensive Strategic Partnership Treaty, ratified in late 2024. The agreement formalises expanded cooperation across scientific and technological domains — including nuclear energy, biotechnology, artificial intelligence, and space development — thereby

creating new channels for knowledge exchange and joint research (U.S. Department of State, 2025).

Persistent economic insecurity among the population is driven by sanctions, limited trade access, and domestic economic mismanagement. At the same time, the regime continues to prioritise investment in strategic weapons programmes, including nuclear, missile, and biological capabilities. This dual strategy preserves regime survival through carefully managed external linkages while sustaining a domestic political order characterised by scarcity, control, and limited transparency (Preble and Willis, 2024). The result is a political economy in which economic hardship among the population coexists with the continued expansion of coercive state power, reinforcing the broader paradox of North Korea as a society that is simultaneously disconnected from global economic integration yet selectively connected through the mechanisms of security competition, sanctions circumvention, and strategic alignment.

Conclusion

It is clear to see that North Korea enables the use of a system in that disconnection is deliberately constructed and strategically maintained. After examining the country with both pillars, it is clear to see that North Korea demonstrates a consistent pattern: external disengagement and internal control are not symptoms of weakness, but instruments of regime survival. Due to the examples provided within these pillars, it is clear to see that the state consolidates authority domestically while selectively engaging with internal systems on its own terms.

This presents a fundamental challenge for policy makers: conventional approaches centered on isolation and pressure risk reinforcing the very dynamics they seek to change. Sanctions and diplomatic engagement, while constraining certain state activities, may inevitably deepen internal disconnection and further entrench regime control. A more effective strategy requires recognising that North Korea’s resilience is rooted in its ability to manage and exploit disconnection.

Naturally, this means that policy responses should prioritise calibrated engagement that targets the societal consequences of this system without strengthening its coercive core.

Ultimately, the objective is not to force rapid transformation, which remains unlikely, but to incrementally reduce the structural depth of disconnection that underpins regime stability. Understanding North Korea as a dis(connected) society reframes the policy challenge: the issue is not simply how to isolate or contain North Korea, but how to engage a system that depends on isolation itself for survival.

Bibliography

Cenciarelli, O., Gabbarini, V., Pietropaoli, S., Malizia, A. and Tamburrini, A. (2013) Biological Weapons and Bioterrorism: A Review of History and Biological Agents. Defence S&T Technical Bulletin. Available at: https://www.researchgate.net/publication/264482643_Biological_Weapons_and_BioTerrorism_a_review_of_Hystory_and_Biological_Agents

Council of Europe (2024) Transnational Repression and Threats Against Scholars. Available at: https://rm.coe.int/policy-brief-18-2/4880283eb7

Freedom House (2021) Out of Sight, Not Out of Reach: The Global Scale and Scope of Transnational Repression. Available at: https://freedomhouse.org/report/transnational-repression

Haggard, S. and Noland, M. (2009) Repression and Punishment in North Korea: Survey Evidence of Prison Camp Experiences. Available at: https://mpra.ub.uni-muenchen.de/17705/

Hwang, K. (2024) Possible bioterrorism by North Korea and South Korea's preparedness, Infection & Chemotherapy, 56(3), pp. 300–307. Available at: https://doi.org/10.3947/ic.2024.0068

Human Rights Watch (2025) North Korea: Escapees Describe COVID-linked Repression. Available at: https://www.hrw.org/news/2025/03/05/north-korea-escapees-describe-covid-linked-repression

Michaelsen, M. and Thumfart, J. (2023) 'Drawing a line: Digital transnational repression against political exiles and host state sovereignty', European Journal of International Security. Available at: https://www.cambridge.org/

Park, P.J. and Glaser, A. (2024) Estimating Potential Tritium and Plutonium Production in North Korea's Experimental Light Water Reactor. Available at: https://arxiv.org/abs/2412.11993

Preble, K.A. and Willis, C.N. (2024) ‘Trading with pariahs: North Korean sanctions and the challenge of weaponized interdependence’, Global Studies Quarterly, 4(2), ksae031. Available at: https://doi.org/10.1093/isagsq/ksae031

Reuters (2017) North Korean leader’s half-brother killed with VX nerve agent in Malaysia. Available at: https://www.reuters.com/article/world/chemical-weapon-vx-nerve-agent-killed-north-korean-leaders-half-brother-malays-idUSKBN16303Z

Stimson Center (2023) Biological Weapons Proliferation. Available at: https://www.stimson.org/wp-content/files/file-
Tsourapas, G. (2021) ‘Global autocracies: Strategies of transnational repression, legitimation, and co-optation’, International Studies Review, 23(3).
U.S. Department of State (2024) Country Reports on Human Rights Practices: North Korea. Available at: https://www.state.gov/reports/2024-country-reports-on-human-rights-practices/north-korea

U.S. Department of State (2025) Adherence to and compliance with arms control, nonproliferation, and disarmament agreements. Available at: https://www.rfa.org/english/korea/2025/04/21/north-korea-biological-weapon/

U.S. Department of the Treasury (2024) Treasury sanctions actors financing North Korea weapons of mass destruction program. Available at: https://home.treasury.gov/news/press-releases/jy2215

Reforming the Role of AI in Compulsory Education

Ethan Higgins

Abstract

This proposal will assess the challenges and opportunities presented by Artificial Intelligence in education for 4–18-year-olds. This essay will analyse the impacts of AI on education on two fronts: firstly, assessing the potential benefits of using AI in teaching-to-learning outcomes and the use cases of Generative Artificial Intelligence for teachers, while balancing this against the challenges posed by AI, such as ethics, accountability, transparency, algorithmic prejudice and security; secondly, it will discuss the imperative nature of teaching students about a world dominated by AI, helping students understand the benefits and drawbacks of AI, while predominantly teaching students to distinguish AI-generated content from authentic content, and teaching them the critical role of trust in a post-truth world. This is an issue which should be addressed holistically in the UK to prepare the next generation for the rapidly changing digital world. To achieve this, this proposal will recommend: the development and implementation of a national AI model for students and teachers; expert training in AI usage for all teachers and training in AI ethics; and risks for school leaders and ICT teachers to allow the topics to be taught to all students.

AI – A Tectonic Shift

Artificial Intelligence can be considered the most impactful, and certainly the most controversial, technology of our time. Nobel Laureate Geoffrey Hinton has said that AI will be "the next Industrial Revolution" (Time Magazine, 2023), bringing with it enormous gains to productivity and earthshaking changes to how we lead our lives. In the world of education, AI presents great opportunities alongside a myriad of ethical and practical considerations. 66.5% of young people aged 13-18 reported the use of Generative AI (GenAI) in their studies, with 45.6% using it weekly or more often. 58% of teachers reported use of AI, and 66.9% said they need more training and support to utilise AI effectively (National Literacy Trust, 2025). Current use of AI in schools, in which students are gently discouraged from using AI and not taught effectively how to use it, has led to 62% of students reporting that AI has had a negative impact on their skills and development (Oxford University Press, 2025). It is clear that AI use in schools is widespread, and thus it would be neglectful for educators and policymakers to ignore the benefits of this revolutionary technology.

AI is developing at breakneck speed, and its capabilities and risks are growing every day. AI-generated imagery, video and audio content are becoming increasingly difficult to distinguish from authentic content (Kramer et al., 2025). In a world ruled by social media, already divided by partisan politics and with trust in the establishment degrading, AI disinformation is becoming increasingly prevalent. It is estimated that over 71% of images shared on social media sites are AI-generated (Forbes, 2025), which will be disproportionately viewed by young people who lack the critical thinking skills to distinguish synthetic and authentic content. Students must be taught how to address AI

outside of the classroom and made aware of the various impacts the technology will have on their lives.

This proposal will seek to recommend that the Department of Education not only implement AI in teaching, but also address this emerging technology holistically to prepare the next generation for a post-truth world dominated by GenAI technology. In order to provide a stable foundation for a society living alongside such revolutionary technology, we must integrate AI in classrooms and in the curriculum.

The Current State of AI Policy in the UK

The UK's Labour government has made it clear they intend to support the adoption of AI technologies in many areas of the public sector to increase efficiency and improve project delivery. The government's AI Opportunities Action Plan focuses heavily on implementation in the private sector and civil service, as well as emphasis on the NHS, but it is very brief in its references to AI in education (DSIT, 2025). The foreword by former Secretary of State, Peter Kyle, for Science, Innovation and Technology stated that the government intends to ensure "AI directly benefits working people by improving health care, education and how citizens interact with their government", but beyond this there is little mention of compulsory education.

The report emphasises the importance of "training, attracting and retaining the next generation of AI scientists and founders". This might be an effective solution, but it reveals the government's lack of commitment to holistic AI education. They appear to be solely focused on education to improve our AI-building capabilities at the expense of our understanding of AI as a society. As a result, the plan focuses primarily on higher education because it appears the government's main priority is teaching how to build AI and digital literacy, not utilising it in classrooms or addressing its societal impacts.

The Benefits of AI in Classrooms

A study by the Department of Education into use cases of GenAI for teachers depicts positive results (DfE, 2024). The research team developed a Proof of Concept GenAI tool and brought together over 60 teachers, education policy experts and data scientists to test the

concept. They found five use cases for teachers with "potential for development". These use cases were scored by participants on five factors: usability, time saved, meeting national standards, improving outcomes, and likelihood to use.

The use cases studied and their results are as follows:

- Generating Age-Appropriate Lesson Plans - Scored well for time saving (3.7/5) but results were weaker for the other four factors, notably likelihood to use (2.3/5) significantly below the average across all use cases (3.6/5). The results suggest that a model could be implemented effectively if functionality is improved to resolve the drawbacks.

- Generating Effective Questioning Ideas – Scored above the average across all cases for all factors except for likelihood to use, particularly for time saved (4.7/5) and meeting national standards (4/5). An easier to use model would save teachers significant amounts of time and would not sacrifice quality in the process.

- Generating Lesson Materials – Greatly exceeded the average across all cases for all five factors, including a perfect score for improving outcomes (5/5). The technology is already in place for effective implementation in this case, with the model scoring highly for likelihood to use (4.5/5).

- Marking Student Work – Received a low score for actual usability (1.8/5) but a high score on likelihood to use if improved (3.8). This would save teachers lots of time if effective. A study outlined in the AI Opportunities Action Plan shows that a more recent "generative AI and rules-based marking tool showed 92% accuracy in a pilot" when marking literacy work for year 4 students, with the tool given greater access to appropriate educational data. (DSIT, 2025). A more accurate model, as is described in the AI Opportunities Action Plan, would resolve the usability issues shown in DfE's results.

- Generating Personalised Feedback for Students – Received below average scores across the board, particularly on likelihood to use (2.3/5). The participants expressed doubts that the model, even if improved, could achieve the accuracy level desired by teachers.

It is clear from the use cases outlined that the potential benefits of AI in classrooms, particularly in learning outcomes for students and saving time for teachers, are enormous.

The benefit of better learning outcomes is self-explanatory, but the benefit of saving time for teachers is perhaps more significant. Reducing the workload of easily automated tasks such as marking and producing lesson materials for teachers would allow them to spend more time personalising lessons and feedback for students, as well as more time to act in a pastoral capacity for students. This would benefit students academically as they would receive better quality feedback to improve their work, but would also allow teachers to address issues which are rampant in schools displaying student loneliness and poor mental health.

In 2025, an NHS study showed that 20.3% of children aged 8-16 showed signs of a "probable mental disorder", displaying characteristics of one or more mental health issues (NHS, 2023). Mental health charity MIND found that 78% of surveyed young people reported their mental wellbeing was negatively impacted by their time at school (MIND, 2021). Relieving pressure on teachers by implementing AI tools would give them the opportunity to spend more time supporting students emotionally, helping to reduce the astronomical figure of unhappy students in a national mental health crisis.

AI could also support early identification of Special Educational Needs (SEN), and may be very effective at modifying and adapting learning resources to support SEN students (Khan & Sujatha, 2025). Implementation of AI in schools has the capacity to benefit SEN students a great deal, as students often spend years in school undiagnosed, and even once diagnosed, are left behind by a collapsing, under-resourced system (Education Committee, 2025). AI tools would allow much cheaper, more effective personalisation for students with varying needs and disabilities, helping relieve pressure from the flailing SEND system in the UK, as well as providing more time for teachers to work closely with students with additional needs.

The Imperative Necessity of AI in the Curriculum

There are reasonable concerns often mentioned when discussing AI, such as issues over a decrease in critical thinking and AI overdependence. One third of students interviewed by Oxford University Press said they could not tell whether content generated by AI chatbots for homework was true, with a further fifth being unsure on its quality (Oxford University Press, 2025). This is especially important; the headteacher of a school involved in the study notes, "many [students] are copying and pasting from AI". This issue is widespread in

society as AI becomes more prevalent and the circulation of AI-generated media becomes increasingly common. In a world of "post-truth" politics, the presence of tools which can be used to produce near-undetectable disinformation (Kramer at al., 2025) is driving people apart and further entrenching a culture of lies. When implementing lessons on AI into the curriculum, teaching students how to investigate the information AI chatbots draw from and how to distinguish AI-generated media from authentic man-made media, is essential, and would better prepare young people for the uncertain future of a world lived alongside AI.

Wellbeing in a world of AI is an issue that will also be tackled from the bottom-up through our education system. The lack of effective teaching around the use of AI and its impact on society is causing increased stress for young people. A study into higher education student wellbeing with AI has found that 68% have reported increased personal stress as a result of insufficient direction from educational institutions around the use of AI (Higher Education Policy Institute, 2025). While this study focuses on higher education, it is relevant here as increased AI education in schools would negate this problem for universities. Students in the study noted they were left to work out how best to use AI tools for themselves, leading to 61% saying they feel moderately less confident they are using AI in an effective way to benefit their education and personal development. Effective teaching in the curriculum on how best to use AI for learning and skills development would make young people much more confident in their studies, and would relieve some of the already significant strain on student mental health.

The other primary concern about AI stems from the ethical concerns and consequences of its usage. All AI tools are biased by virtue of their development. As a result of the differing data used to train AI tools, different chatbots and image-generation software will give different results when prompted in the same way (Akgun & Greenhow, 2021). Often, these only show slight semantic differences, or somewhat differing sources, but AI can be developed with intentional bias within it. A commonly used example of this is Grok, a chatbot integrated in the social media platform X and championed by Elon Musk. Investigation by the UK fact-checking organisation, Full Fact, found that Grok is often unreliable in the veracity of its information, and often "seemed to promote Musk's favoured right-wing talking points" (The Guardian, 2025). AI can also be unintentionally biased through what is known as "Algorithmic Prejudice", where, due to the enormous quantity of man-made data used in training AI, biases inherent within society, such as racism and sexism, are disproportionately

present in AI generated content (Cataleta, 2020). Entirely eliminating these biases is improbable, but teaching students of these biases throughout their time at school would help reduce its impact and increase students' critical thinking abilities.

A Blueprint for Success

Establishing AI in both teaching and learning is not an easy process, and is one which would take a considerable amount of time. Below outlines potential short-term and long-term plans for how to transition to a system which embraces AI.

The most immediately achievable goal for the DfE is to increase training for teachers and school leaders in the use of AI in their work. A one-off crash course in AI education tools for school leaders (approximately 52,000 people), taught preferably over the summer break for 35 hours per head at a cost of £80 per hour would prepare schools for the AI transition (Education Policy Institute, 2021). This would result in a setup cost of £200 million for the program. It would also require an extension of Continuing Professional Development (CPD) for all teachers by 3 hours per year to maintain understanding of use of the tools, as well as redeveloping the CPD scheme to focus more on digital and AI training (Education Policy Institute, 2021). This would cost £160 million per year as an ongoing cost, and be relatively easy to implement as it uses existing systems for teacher training. ICT teachers would also require a further 3 hours of extra CPD training per year to implement AI in the curriculum, with members of staff at all schools being aware of the benefits and risks, as explained in previous sections.

The training would focus on synthetic content detection, algorithmic prejudice and critical thinking to allow the necessary skills and understanding to be passed on to students. This would cost another £160 million annually. The program is limited here, as it would require introducing ICT teachers to the 31% of schools which currently do not offer the program at A-level (Teach First, 2025). One way to combat this limitation would be to create incentives for trainees to become ICT teachers by increasing their salary by 5% for the first five years of their teaching, which would ease the supply issues at a cost of £37 million annually (Sims, 2017). This brings the total cost for human training to roughly £360 million per year. The DfE currently spends just over £1 billion annually on CPD training, so this would require significant increases in funding for the department.

The larger and less immediately achievable objective for transitioning to AI-supported education is implementing the necessary digital infrastructure both at the DfE, and in schools across the country. The first step to achieve this objective is for the government to develop their own in-house AI model for education. A model built by the government is preferable to the use of private models as it allows complete control over the data used for training, allowing the government to minimise algorithmic prejudice and avoid ideological bias in teaching. This is estimated to cost the UK government £40 million in setup costs (half of the total cost of OpenAI building GPT-4). There would also be ongoing costs for keeping the model up-to-date although there is no available estimate on that figure. The next stage of the process would be to implement Digital Learner IDs for all students. This data system would hold all digital data for students and would be platformed on a unified online portal accessible by schools, students and parents. This is estimated to cost roughly £50 million in setup costs and a further £10 million annually to maintain (Browne et al., 2024). The final, and most expensive, step would be to provide all schools with the necessary technology to implement use of the government AI edtech model for all students. The Tony Blair Institute estimates this would require the procurement and distribution of 6 million high-tech tablets across the nation. This would cost approximately £600 million year-on-year to purchase and maintain.

The cost of this proposal is significant and limits its practicality, however there is room for the spending in the Department's budget. The total £1 billion needed for the program could be funded by the recent introduction of VAT on private schools, which is expected to raise £1.51 billion in the 2025/26 school year to be spent on education (Seely, 2025).

The Future of AI-Adjacent Education

In total for the program, the initial setup cost is £290 million. This is an achievable figure for the department which has a yearly budget of over £100 billion. The real test of funding is in the ongoing costs, which total approximately £1 billion per year, assuming a cost under £40 million per year of maintaining the government's AI model. There would be some savings for councils' SEN budgets with the implementation of AI, although there is no estimate on this figure yet. This is certainly an expensive project for the department, but the benefits to education quality and developing new life skills for students in a digital world, as well as potential benefits to student wellbeing, outweigh the costs.

The implementation of AI tools in schools, while expensive, can be a fast-moving process. If action is taken now, students can be empowered by AI tools and expert teaching in AI ethics, accountability and security by 2030. There are faster alternative methods of implementing AI in schools which may be considered. Licensing private AI tools like ChatGPT to serve in place of a government-run education model would cut down on the time needed to set up the program, but there are multiple reasons this is less desirable for the government. For one, building a national AI model is preferable to licensing use of private models as it avoids potential catastrophes involving data protection and ethical usage, and access to private AI tools in all schools would cost £400 million per year, far more than developing a purpose-built, nationally run model.

Other countries which have implemented AI in schools have taken alternative paths. South Korea has introduced AI tools in classrooms and integrated them into their curriculum, but their focus has been on AI building to boost technical proficiency, problem solving and self-directed learning to drive economic growth and enhance global competitiveness in the AI market. This ignores the critical issues of AI ethics and AI-adjacent skill development, which this article suggests are essential to effective AI education. Japan has implemented AI in schools and their curriculum in a similar way to the recommendations made in this article, highlighting integrative thinking, critical and innovative problem-solving, and creativity in what is described as a "society-centred" approach (Shin-Hye, 2025). However, they have opted to use private models instead of their own purpose-built model.

It is important to note that measuring the success of this proposal will be difficult. Whilst benefits to learning outcomes of students currently in secondary school would be clearly shown in increased attainment, the full effects of the policy on children using AI from primary through to GCSE/A-Levels would take approximately a decade to measure effectively, as students would be trained in using AI from childhood, which means its benefits may not seem as impactful in the short-term. Evaluating the success of AI in the curriculum is even more opaque, as there is no practical measure for critical thinking. Studies investigating people's abilities to distinguish synthetic and authentic media would perhaps show a slight positive change following the implementation of this policy but, at the rate GenAI models are developing, it is likely this would go unnoticed.

Bibliography

Akgun, Selin, and Christine Greenhow. “Artificial Intelligence in Education: Addressing Ethical Challenges in K-12 Settings.” *AI and Ethics*, vol. 2, no. 3, 22 Sept. 2022, pp. 431–440, www.ncbi.nlm.nih.gov/pmc/articles/PMC8455229/, https://doi.org/10.1007/s43681-021- 00096-7.

Booth, Robert. “In Grok We Don’t Trust: Academics Assess Elon Musk’s AI-Powered Encyclopedia.” *The Guardian*, The Guardian, 3 Nov. 2025, www.theguardian.com/technology/2025/nov/03/grokipedia-academics-assess-elon-musk-ai powered-encyclopedia.

Browne, James, et al. *The Economic Case for AI-Enabled Education a Companion to the Economic Case for Reimagining the State*. 2024.

Cataleta, Maria. “Humane Artificial Intelligence: The Fragility of Human Rights Facing AI the Fragility of Human Rights Facing AI on JSTOR.” *Jstor.org*, 2024, www.jstor.org/stable/resrep25514?searchText=Ethical+AI&searchUri=%2Faction%2FdoBasi cSearch%3FQuery%3DEthical%2BAI%26so%3Drel&ab_segments=0%2Fbasic_search_gsv 2%2Fcontrol&refreqid=fastly-default%3A1ad33e0c0ab5529b8b4f66f9dc5e4fd6, https://doi.org/10.2307/resrep25514.

Department of Education. *Use Cases for Generative AI in Education User Research Report*. 2024.

GOV.UK. “AI Opportunities Action Plan.” *Gov.uk*, 13 Jan. 2025, www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action plan. Accessed 10 Mar. 2026.

Hinton, Geoffrey. “AI Is the next Industrial Revolution.” *TIME*, Time, 11 Dec. 2025, time.com/7339628/geoffrey-hinton-ai/.

Kelly, Joni. *Edtech Decision-Making and Inclusive Practice: Insights from Multi-Academy Trusts*. Education Policy Institute, 2025.

Khan, Reshma, and S. Sujatha. “AI in Special Education: Enhancing Learning Outcomes for Students with Disabilities.” *International Journal of Accounting and Economics Studies*, vol. 12, no. SI-1, 28 Aug. 2025, pp. 97–105, www.researchgate.net/publication/395050749_AI_in_Special_Education_Enhancing_Learni n g_Outcomes_for_Students_with_Disabilities#citations, https://doi.org/10.14419/ana2t992.

Kramer, Robin S. S., et al. “AI-Generated Images of Familiar Faces Are Indistinguishable from Real Photographs.” *Cognitive Research: Principles and Implications*, vol. 10, no. 1, 14 Oct. 2025, https://doi.org/10.1186/s41235-025-00683-w.

Marr, Bernard. “15 Mind-Blowing AI Statistics Everyone Must Know about Now.” *Forbes*, 10 Mar. 2025, www.forbes.com/sites/bernardmarr/2025/03/10/15-mind-blowing-ai-statistics everyone-must-know-about-now/.

Mind. “Not Making the Grade: Why Our Approach to Mental Health at Secondary School Is Failing Young People.” 2021.

National Literacy Trust. “Young People and Teachers’ Use of Generative AI to Support Literacy in 2025.” *National Literacy Trust*, 28 Aug. 2025, literacytrust.org.uk/research services/research-reports/young-people-and-teachers-use-of-generative-ai-to-support-literac y in-2025/.

Nazaretsky, Tanya, et al. “Teachers’ Trust in AI-Powered Educational Technology and a Professional Development Program to Improve It.” *British Journal of Educational Technology*, vol. 53, no. 4, 9 May 2022, bera-journals.onlinelibrary.wiley.com/doi/10.1111/bjet.13232, https://doi.org/10.1111/bjet.13232.

---. “The Critical Role of Trust in Adopting AI-Powered Educational Technology for

Learning: An Instrument for Measuring Student Perceptions." *Computers and Education: Artificial Intelligence*, vol. 8, June 2025, p. 100368, https://doi.org/10.1016/j.caeai.2025.100368.

NHS Digital. "Mental Health of Children and Young People in England, 2023 - Wave 4 Follow up to the 2017 Survey." *NHS Digital*, NHS England, 2023, digital.nhs.uk/data-and information/publications/statistical/mental-health-of-children-and-young-people-in england/2023-wave-4-follow-up. 21 November 2023.

Sam, Sims. What Happens When You Pay Shortage-Subject Teachers More? The Gatsby Foundation, 2017.

Scales, James, et al. "Generation Ready: Scaling Safe, High-Quality AI in England's Schools." Tony Blair Institute, 22 Jan. 2026.

Seely, Antony. "VAT on Private School Fees." *House of Commons Library*, 27 June 2025, commonslibrary.parliament.uk/research-briefings/cbp-10125/.

Shin-Hye, Heo. "Assumptions on Core Competencies in AI Education: A Comparative Policy Analysis of South Korea and Japan." *International Journal of Advanced Smart Convergence*, vol. 14, no. 3, 2025, pp. 306–315, koreascience.kr/article/JAKO202529639643490.page, https://doi.org/10.7236/IJASC.2025.14.3.306. Accessed 10 Mar. 2026.

"Teacher Shortages Deny Nearly a Third of Poorer Pupils' Access to Computer Science A Level | Teach First." *Teachfirst.org.uk*, 8 May 2025, www.teachfirst.org.uk/press release/missing-teachers.

Teaching the AI-Native Generation Empowering Schools in the Age of AI. Oxford University Press, Aug. 2025.

The Open Innovation Team and Department for Education. *Generative AI in Education Educator and Expert Views Authors*. Jan. 2024.

UNESCO. *Working Papers on Education Policy 07 Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development Education Sector United Nations Educational, Scientific and Cultural Organization*. 2019.

Van den Brande, Jens, and James Zuccollo. *The Effects of High-Quality Professional Development on Teachers and Students: A Cost-Benefit Analysis*. Education Policy

Institute, Apr. 2021.

YouGov. *Global Survey - HE Student Wellbeing* . YouGov, 2025.

The Paradox of Connection: A Behavioural Game-Theoretic Analysis of Social Distance and Public Policy

Joe Vick

Abstract

This article develops a behavioural game-theoretic model to analyse how social distance shapes societal welfare. It extends on Fiedler et al. 's (2011) work by presenting the trade-off between perceived trust and potential gains from virtual interaction as a continuous and concave function. The model predicts that welfare is maximised at an intermediate level of social distance. It is applied to three areas of UK public policy: digital public services, social media infrastructure, and labour market flexibility. Drawing on empirical evidence, the article shows that recent policy trends have prioritised efficiency and connectivity at the expense of trust, leading to suboptimal outcomes in welfare. The article then uses the model to propose reforms that could rebalance this trade-off and improve welfare. Overall, the findings show the value of incorporating behavioural insights into modern public policy design.

Introduction

Modern societies can be described as more socially connected than ever before. Rapid expansions in digital communication technology, global markets and online public services have played a major role in integrating individuals with the wider population. In the United Kingdom, the government has supported these developments by digitalising services and promoting online platforms and flexible labour markets. These policy decisions are primarily guided by the HM Treasury Green Book, which uses a cost-benefit framework that associates increased connection with efficiency and productivity gains (HM Treasury, 2026).

However, empirical evidence on social connection points in the opposite direction. Across 21 European OECD countries, the share of people meeting friends daily fell from 21% in 2006 to 12% in 2022, while the proportion contacting friends remotely rose over this period (OECD, 2025). At the same time, indicators of loneliness have worsened. Among young people, the share reporting they "never feel lonely" fell by 14% between 2018 and 2022 (OECD, 2025). Together, these trends suggest that increased connectivity has not translated into increased human interaction, and may even be contributing to its decline.

This raises a central question: why have policies designed to increase connectivity not produced stronger social cohesion? This article argues that government policies in the UK which aim to increase connection can unintentionally exacerbate feelings of social distance, defined as the perceived distance between individuals or groups (Dufwenberg and Muren, 2006). Using evidence from trust game experiments (Berg et al., 1995; Fiedler et al., 2011), this article shows that individuals are less willing to trust partners they perceive as more socially distant. A simple model is then developed linking social distance with an individual's inclination to cooperate and their potential payoff from doing so. Interestingly, when perceived distance approaches either extreme (very close or very distant), interactions produce less welfare for the parties involved. Using this framework, the article evaluates three policy areas in the UK: digital government services, social media infrastructure, and platform labour markets. This will demonstrate how policies justified primarily through efficiency and connectivity may unintentionally weaken social cohesion through increased polarisation or alienation.

Model

In this section, a model is constructed to examine how changes in social distance influence individual welfare. A useful starting point is the trust game experiment introduced by Berg et al. (1995). In this game, two players participate with no prior communication: the Proposer (A) and the Responder (B). Both players receive an endowment of $10. The proposer chooses how much of this endowment to send to the responder, denoted by x. The amount sent is tripled when it reaches the responder, meaning they receive $3x$. The responder then decides how much to return to the proposer, denoted by y. The resulting monetary payoffs for the two players are (Berg et al., 1995):

$$\pi_A = 10 - x + y$$

$$\pi_B = 10 + 3x - y$$

The results of the experiment show that an average of $5.16 was sent to the responder, with 30 out of 32 proposers sending a positive amount. Of the responders who received more than $1, 11 out of 28 returned more than they had been sent. These outcomes show that individuals often exhibit trust and reciprocity when making cooperative decisions, even without prior interaction.

Fiedler et al. (2011) extend this trust game to examine how social distance affects trust in online environments. Using the virtual world 'Second Life' and its currency, Linden Dollars, proposers (A) are presented with two potential responders. One responder is located in the same virtual room as the proposer (the in-room responder, B) and can communicate with them beforehand through text chat and avatar interaction. The second responder is located in a separate room (the out-room responder, C) and cannot communicate with the proposer. Communication is expected to reduce perceived social distance by allowing the proposer to gather information about the responder's behaviour. All participants start with 1000 Linden Dollars. The proposer must first choose which responder to play the trust game with, and then decide how much of their 1000 Linden Dollar endowment to send. The amount sent is again denoted by x, and the amount returned by y. However, the multiplier (m) applied to x differs depending on the partner chosen. If the proposer selects B, the multiplier is 2. If C is selected, the multiplier is either 3 or 4.

To prevent feelings of guilt influencing the proposer's decision, the experimental design also gives B the same returned amount y whenever the proposer chooses C. The resulting payoffs are therefore (Fiedler et al., 2011):

If the proposer chooses B:

$$\pi_A = 1000 - x_B + y_B$$
$$\pi_B = 1000 + m_B x_B - y_B$$
$$\pi_C = 1000$$

If the proposer chooses C:

$$\pi_A = 1000 - x_C + y_C$$
$$\pi_B = 1000 + y_c$$
$$\pi_C = 1000 + m_C x_C - y_c$$

Choosing the socially distant partner therefore creates a higher potential monetary return, as C receives a larger multiplier on the money sent, and B also gains y.

Responders are also assumed to differ in their preference for reciprocity, where H-type responders are considered more inequality-averse than L-type responders. As a result, H -types return larger amounts compared to L-types because returning more money produces a more equal division of the final payoffs. The utility of the responders can therefore be written as (Fiedler et al., 2011):

$$U_j^T = \pi_j - \beta^T \frac{(\pi_j - \pi_A)^2}{x_j}$$

where $j = \{B, C\}$ and $T = \{H, L\}$. In this generalised formulation, a responder's utility is their own payoff minus the disutility created by unequal outcomes with the proposer. Higher values of β represent stronger inequality aversion. Solving this expression for maximum utility gives: (see appendix)

$$y_j^{T*} = \frac{1}{2}\left(m_j + 1 - \frac{1}{4\beta^T}\right) \times x_j$$

which represents the amount a responder should return when acting in their own best interest. For example, if the proposer sends 500 Linden Dollars to a high-trust responder B and $\beta^H = 0.8$, the utility-maximising response is:

$$y_B^{H*} = \frac{1}{2}\left(2 + 1 - \frac{1}{4(0.8)}\right) \times 500 = 671.875\ Linden\ Dollars$$

The proposer cannot know whether a responder is a *H*- or *L*-trust type, but forms beliefs about where they are more concentrated. Let q represent the probability that B is a *H*- type, and p the probability that C is a *H*-type. Because communication allows for collaborative efforts to feel more secure, the model assumes:

$$q > p$$

The proposer's expected utility therefore depends on the probability of interacting with each type of responder. Assuming the responder behaves in their own best interest, the proposer expects (Fiedler et al., 2011): (see appendix)

$$EU_A^*(j) = f \times (\pi_A(j, H)) + (1 - f) \times (\pi_A(j, L))$$

where $f = \{q, p\}$ and $(\pi_A(j, T))$ is the utility gained when the responder returns y_j^{T*}. In other words, the proposer can expect the weighted sum of each potential payoff they could receive - given the responder is acting in their own best interests - as either a *H*- or *L*-type.

The results of the experiment reflect the model, showing that proposers were more likely to choose the socially closer responder B, doing so 64% of the time when $m = 3$ and 55% when $m = 4$. In addition, B returned a larger proportion of their money on average than C. When $m = 3$, B returned 43% compared with 31% for C, and when $m = 4$, B returned 40%

compared with 32% for C. These findings indicate that individuals favour partners they perceive as socially closer when deciding whether to cooperate.

However, because C receives a higher multiplier and B still receives y when C is chosen, selecting the socially closer partner often produces a lower joint payoff. As shown in Table 1, the out-room game typically generates larger gains for both the proposer and the wider group for either value of m.

Environment	Treatment	Matching Choice	Proposer Payoff	In-Room Responder Payoff	Out-Room Responder Payoff	Joint Payoffs
Second Life	Selection 2-3	Out-room	1299	2007	2117	5423
Second Life	Selection 2-3	In-room	1352	1358	1000	3710
Second Life	Selection 2-4	Out-room	1523	2173	2427	6123
Second Life	Selection 2-4	In-room	1344	1238	1000	3582

Table 1: Payoff comparison (in linden dollars) between in-room and out-room trust games. Source: Fiedler et al. (2011)

This outcome reflects two opposing forces within the model. Increasing social distance raises the potential gains from cooperation, but it also reduces perceived trustworthiness between agents, lowering the probability of these interactions occurring.

The proposer's expected utility therefore varies concavely with respect to perceived social distance, due to the opposing effects that distance has on expected trust f and potential returns m (see appendix). The model also assumes that f declines more rapidly as social distance grows and factors like language barriers and social norms take precedent. Graphically, this relationship can be represented by a curve with each point reflecting the highest expected payoff available at a given level of social distance, shown in Figure 1. This

extension of Fiedler et al. 's model therefore predicts that an intermediate level of distance maximises expected utility.

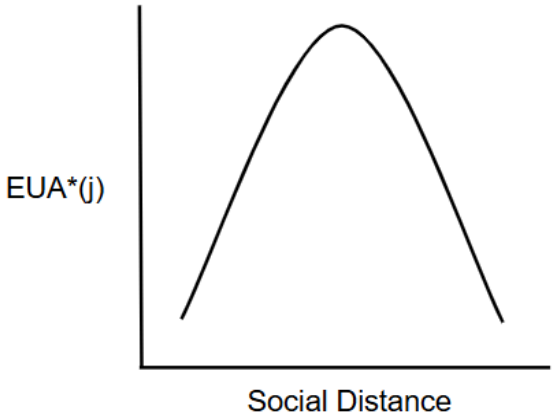


Figure 1

At one extreme, very close social distance limits potential gains because interactions occur only with close partners who offer smaller returns (f is prioritised over m). At the other extreme, very high social distance discourages cooperation altogether, causing individuals to avoid potentially profitable interactions due to risk (m is prioritised over f). In both cases, the outcome is a reduction in realistically achievable utility.

This framework helps answer the question raised in the introduction. Policies designed primarily to maximise connectivity often assume that increasing the number of possible interactions automatically improves welfare. However, the trust game evidence shows that cooperative outcomes are maximised when interactions involve a balance between perceived trust and potential gain, rather than when either factor dominates entirely. Overall, the model presented here follows and extends the experimental framework developed by Fiedler et al., but applies it to the broader question of how perceived social distance online may influence welfare outcomes under different policy environments.

Case Studies

This model can now be applied to various government policies that aim to connect individuals with the rest of society, where welfare becomes limited by the emphasis made on either the potential payoff, or trust. Not both.

Digital Government Services

Over the past decade, the UK government has been increasingly providing services online in order to improve efficiency and reduce administrative costs. One example is unified digital identity systems. The GOV.UK 'One Login' platform allows individuals to access multiple government services through a single account and identity verification process. The system has expanded rapidly since its introduction, with more than 120 different government services now integrated into the platform (Trendall, 2026). These reforms are typically justified within the cost-benefit framework used by the HM Treasury Green Book, where digitalisation is expected to lower transaction costs and streamline service delivery (essentially increasing m in the model). However, this expansion of digital infrastructure also changes the nature of these interactions. By becoming more automated and less socially embedded, individuals increasingly interact with institutional processes rather than other humans.

Within the model developed previously, this shift can be interpreted as a decrease in trust (f) between citizens and public institutions. Processes that previously involved direct contact with an administration are being mediated by technology, which can make institutions appear more distant for a population.

At the same time, the provision of these digital public services has exacerbated forms of exclusion. Estimates suggest that around 7.9 million people lack basic digital skills and 3.7 million households fall below a minimum 'digital living standard', preventing them from effectively engaging with these online services (Good Things Foundation, 2024).

Families on lower incomes and some minority ethnic groups are also significantly more likely to lack the digital resources required to engage with online services. Research shows that people from Black and Asian ethnic backgrounds are less likely to have all essential digital skills compared with those from White backgrounds. This exclusion is also strongly correlated with lower household income (UCL Institute for Global Prosperity, 2020).

This evidence indicates that many UK citizens become less willing and able to rely on administrative systems when interactions move to taking place online, even if it would be more socially efficient to do so (Bélanger and Carter, 2008).

Therefore, by maximising m and neglecting f, the government can be seen to induce a sub-optimal level of welfare. This can be seen in figure 2 where the maximum utility that could be achieved is 'overshot'.

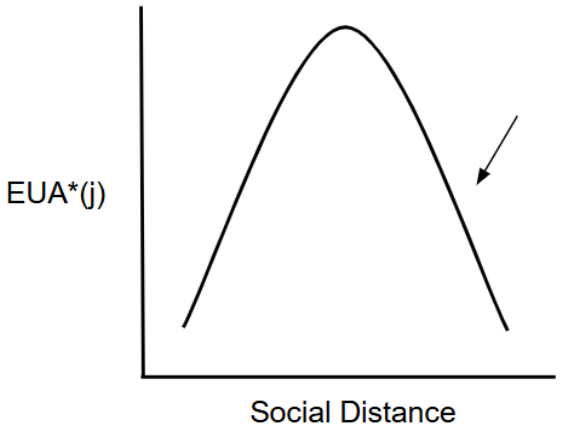


Figure 2

Overall, policies designed to increase connectivity through digital infrastructure may produce a paradoxical outcome: while interactions between citizens and the state become more socially efficient, they also induce feelings of institutional distance and exclusion that limit achievable utility.

Social Media Infrastructure

Government policy has also shaped the structure of the digital information environment through its approach to competition in online markets. For much of the past decade, regulation allowed a highly concentrated digital advertising market to develop. The UK Competition and Markets Authority (CMA) found that around 80% of UK online advertising spending flows to Google and Facebook, indicating a market where a small number of platforms dominate how information is distributed online (Competition and Markets Authority, 2020).

This market structure is important because the dominant platforms rely on advertising revenue. Their systems are therefore designed to maximise user engagement, as greater engagement generates more advertising income. As a result, the information users encounter is often shaped by algorithms that prioritise content most likely to provoke interaction rather than balance exposure to different viewpoints.

Within the model developed previously, this environment significantly reduces perceived trust (f) between social groups. When online information flows are organised around engagement, users are more likely to encounter content produced by people with similar views. Over time, this leads to more fragmented online communities where individuals primarily interact within their own groups.

A potential counterargument here is that social media can foster new forms of community by enabling individuals to connect across geographic boundaries and build relationships around shared interests. However, while these platforms may strengthen connections within groups, they can simultaneously weaken connections between them. Algorithmic curation tends to prioritise content aligned with users' existing preferences, increasing interaction among like-minded individuals and limiting exposure to opposing views. This can reinforce divisions between groups and contribute to greater social fragmentation (Sunstein, 2001).

Government policy has increasingly recognised these problems. The CMA concluded that competition is not working well in digital advertising markets, creating 'substantial harm for consumers and society' (Competition and Markets Authority, 2020). However, these concerns emerged only relatively recently, in which time a small number of platforms were able to accumulate significant market power.

In the context of the model, this reduces welfare. When individuals are rarely exposed to different views, they feel less connected to wider society while becoming increasingly unwilling to interact with those outside their own group who they perceive as less trustworthy. This position can be shown on figure 3, where the high emphasis on maximising f places a limit on m, preventing social distance from reaching its optimal level.

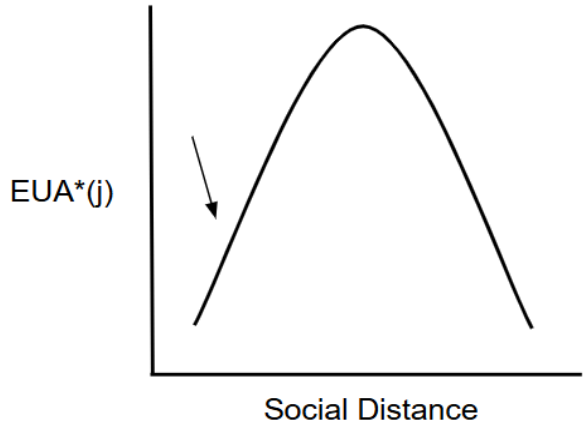


Figure 3

Overall, digital infrastructure that appears to increase connection can simultaneously deepen social distance between groups. The platforms dramatically expand the number of people individuals can communicate with, yet the economic structure of the market (shaped in part by government competition policy) can channel those interactions into increasingly isolated communities and create the growing disconnection perceived today.

Labour Market Flexibility

Government policy has also shaped the labour market through its long-standing emphasis on flexibility, with policy generally supporting self-employment and temporary work and platform labour. These arrangements are often justified as improving labour market access and efficiency by allowing workers and firms to match more easily (Employment Relations (Flexible Working) Act, 2023).

However, these structures also change the social organisation of work. Research suggests that around 4.4 million people in England and Wales undertake gig economy work at least once a week, a figure that has risen sharply since the mid-2010s (TUC and University of Hertfordshire, 2021).

Unlike traditional employment, many forms of platform work involve limited interaction with colleagues or employers. Work is often coordinated online and algorithmically rather than through shared workplaces. This shift matters because workplaces have historically been an important site of social connection outside of the family. Stable employment relationships

allow workers to form friendships, participate in collective institutions such as unions and therefore gain a sense of belonging within the wider society. When work becomes more fragmented and individually mediated through digital platforms, these forms of connection can weaken. Evidence from UK survey data suggests that gig workers experience higher levels of loneliness than those in regular employment, reflecting the more isolated nature of the work (Wang et al., 2022).

Within the model developed in this article, these labour market structures can increase potential gains (m) by making work, and therefore income, more available. However, this is overshadowed by a significant decrease in trust (f) between individuals and the wider economy. Work may remain available, but the social relationships historically embedded within employment become weaker. Figure 2 (used previously) also shows this position, with government policy over-extending connection past what is optimal for welfare and undermining the increased autonomy and income opportunities offered.

In this way, labour market policies designed to increase flexibility can contribute to another form of modern disconnection: eroding the social institutions that once connected workers to the broader economy. The result is a labour market that remains economically active, yet increasingly fragmented and socially disconnected.

Overall, this section has shown that a number of government policies have failed to find a stable balance between their potential gain (m) and the perceived trust (f) that is required for society to participate in them.

Consideration of Reforms

The case studies suggest that the root problem lies in a failure to balance the potential gains from interaction with the trust required for individuals to participate in them. In the model developed earlier, welfare is maximised when perceived social distance lies nearer the middle of the spectrum. Policies that maximise efficiency, connectivity or flexibility above all else risk pushing social distance away from its optimal level. Therefore, effective policy should restore the balance between trust (f) and the potential payoff from interaction (m).

In the case of digital public services, policies that instead provide a hybrid service delivery could help recover the lost institutional trust while retaining some efficiency gains of digital infrastructure. This could include the reintroduction of human intermediaries that provide digital support, such as in-person help centres or staffed telephone services that are equipped to help with more complex processes like benefits claims or tax disputes. These roles would act as a bridge between individuals and digital systems, allowing users to receive guidance and reassurance throughout these processes. As a result, the model predicts that such policies would increase perceived trust (f) in public institutions, a key determinant of adoption (Bélanger and Carter, 2008). At the same time, because the underlying infrastructure remains digital, many of the efficiency gains (m) associated with standard and automatic processes that digitalisation was designed to achieve could still be preserved to a significant extent, potentially counteracting the higher administrative costs of hiring these human intermediaries. Within the model, this shifts interactions closer to the welfare-maximising level of social distance (see Figure 1).

Social media infrastructure could also be improved through policies that increase competition and regulate how content is recommended, thereby widening the range of information users encounter. This could include measures such as enforcing higher transparency in recommendation algorithms, or mandating that a proportion of recommended content is drawn from outside a user's preference profile to ensure exposure to a wider range of viewpoints, limiting purely engagement-driven content sorting. These interventions would weaken the dominance of a small number of platforms and reduce the extent to which users are channelled into echo chambers. As a result, individuals would be more likely to encounter a broader range of viewpoints, increasing perceived trust (f) between social groups by reducing informational asymmetries and reinforcing shared understanding. At the same time, because users would still benefit from large-scale digital networks, many of the gains from connectivity (m) could be preserved, although stricter regulation may reduce engagement and advertising efficiency. Within the model, this would shift social distance away from artificially low levels and closer to the welfare-maximising level shown in Figure 1.

Finally, platform labour and highly flexible employment arrangements could be reformed through policies that increase worker stability. This could include measures such as extending union representation and, therefore, collective bargaining rights to platform workers, or

requiring minimum-hour guarantees to reduce uncertainty in income and working conditions. These interventions would partially reintroduce the stable social structures historically associated with employment, allowing workers to form more consistent relationships with employers and colleagues. These stronger connections within the labour market would increase perceived trust (f) in economic institutions and among participants. At the same time, because platform-based matching and digital coordination would remain in place, many of the efficiency gains (m) associated with flexible labour markets could be preserved, although such reforms may reduce flexibility for firms and increase labour costs. Within the model, this would shift interactions away from excessively high levels of social distance and closer to the welfare-maximising level shown in Figure 1.

Taken together, these examples highlight a broader lesson for policymakers. Economic systems can increase connectivity in a purely technical sense while simultaneously discouraging forms of social connection that maximise the benefits of said connection. Policies cannot afford to only prioritise one variable in the model, with a more effective approach being to design institutions that maintain the balance between trust and potential gain that underpins cooperation in the first place.

Conclusion

This article has used empirical evidence from behavioural game-theoretic experiments to formulate a model that accurately relates an individual's utility to their degree of social distance between others, finding that an intermediate distance is able to capture both the payoff and trust required to maximise utility. By applying the model more broadly to various government policies that aim to maximise societal welfare through increasing digital connectivity, a number of flaws can be found within the reasoning behind them, where they can be seen to neglect at least one variable in the model and, overall, dampen their potential positive effects.

Bibliography

Bélanger, F. and Carter, L. (2008) 'Trust and risk in e-government adoption', *Journal of Strategic Information Systems*, 17(2), pp. 165-176.

Berg, J., Dickhaut, J. and McCabe, K. (1995) 'Trust, reciprocity and social history', *Games and Economic Behavior*, 10(1), pp. 122-142.

Competition and Markets Authority (2020) 'Online platforms and digital advertising: market study final report'. *London: CMA*.

Dufwenberg, M. and Muren, A. (2006) 'Generosity, anonymity and social distance in dictator games', *Journal of Economic Behavior and Organization*, 60(1), pp. 42-55.

Employment Relations (Flexible Working) Act 2023. London: UK Parliament.

Fiedler, M., Haruvy, E. and Li, S. (2011) 'Social distance in a virtual world experiment', *Games and Economic Behavior*, 72(2), pp. 400-426.

Good Things Foundation (2024) *Digital exclusion in the UK 2024*. Available at: https://www.goodthingsfoundation.org (Accessed: 15 March 2026).

HM Treasury (2024) The Green Book: *UK Government Guidance on Appraisal*, London: HM Treasury, pp. 13

OECD (2025) *Social connections and loneliness in OECD countries*. Paris: Organisation for Economic Co-operation and Development.

Sunstein, C. (2018) *Republic: Divided Democracy in the Age of Social Media*. NED-New edition, Princeton University Press

Trendall, S., (2026) 'GOV.UK One Login expands to over 120 services'. *PublicTechnology*. Available at: https://www.publictechnology.net (Accessed: 15 March 2026).

Trades Union Congress and University of Hertfordshire (2021) *Gig economy in the UK: scale and characteristics of platform work*. London: TUC.

UCL Institute for Global Prosperity (2020) *Digital inclusion and inequality in the UK*. London: University College London.

Wang, B., Liu, Y., Qian, J. and Parker, S. (2022) 'Achieving effective remote working during COVID-19', *Applied Psychology*, 71(1), pp. 16-59.

Appendix A: Derivation of the Responder's Optimal Return

The monetary payoffs are:

$$\pi_A = 1000 - x_j + y$$
$$\pi_j = 1000 + m_j x_j - y$$

Substituting these payoffs into the responder utility function:

$$U_j^T = (1000 + m_j x_j - y) - \frac{\beta^T}{x_j}[(1000 + m_j x_j - y) - (1000 - x_j + y)]^2$$

Simplifying:

$$U_j^T = 1000 + m_j x_j - y - \frac{\beta^T}{x_j}[(m_j + 1)x_j - 2y]^2$$

Differentiating with respect to y gives:

$$\frac{\delta U_j^T}{\delta y} = -1 + \frac{4\beta^T}{x_j}[(m_j + 1)x_j - 2y]$$

Solving the first-order condition gives:

$$-1 + \frac{4\beta^T}{x_j}[(m_j + 1)x_j - 2y] = 0$$
$$2y = (m_j + 1)x_j - \frac{x_j}{4\beta^T}$$
$$y = \frac{1}{2}[(m_j + 1) - \frac{1}{4\beta^T}]x_j = y_j^{T*}$$

Appendix B: Expected Utility of the Proposer

Proposer (A)'s payoff as stated in the main text:

$$EU_A^*(j) = f \times (\pi_A(j,H)) + (1-f) \times (\pi_A(j,L))$$

Where $\pi_A(j,H)$ represents A's payoff

Substituting y_j^{T*} into $\pi_A(j,T)$ and simplifying gives:

$$\pi_A(j,T) = 1000 + [\frac{1}{2}(m_j + 1 - \frac{1}{4\beta^T}) - 1]x_j$$

Substituting $\pi_A(j,T)$ into $EU_A^*(j)$ gives:

$$EU_A^*(j) = f[1000 + (\frac{1}{2}(m_j + 1 - \frac{1}{4\beta^H}) - 1])x_j] + (1-f)[1000 + (\frac{1}{2}(m_j + 1 - \frac{1}{4\beta^L}) - 1)x_j]$$

This expression shows that the proposer's expected payoff depends on:

A. The multiplier m_j associated with the chosen responder
B. The level of inequality aversion among responders
C. The proposer's belief f that the responder will reciprocate generously

Higher perceived social proximity increases f, raising the expected payoff from cooperative interactions even when the potential monetary multiplier is lower

Appendix C: Proof of Concavity

By collecting like-terms in the proposer's expected utility (appendix B):

$$EU_A^*(j) = 1000 + (\frac{m_j - 1}{2})x_j - \frac{x_j}{8}(\frac{f}{\beta^H} + \frac{1-f}{\beta^L})$$

In the experiment described by Fiedler et al. (2011), both of these variables depend on perceived social distance, denoted here as D. Greater social distance increases the multiplier but reduces perceived trust. Thus we write: (note: this model generalises the discrete values of m in Fiedler et al. into a continuous function $m(D)$ to represent the broader range of payoffs present in society. $f(D)$ represents the decay of q towards p as social distance increases)

$$m_j = m_j(D), \quad f = f(D)$$

With:

$$m_j'(D) > 0, \quad f'(D) < 0$$

Substituting these values into the expected utility function gives:

$$EU_A^*(D) = 1000 + \left(\frac{m_j(D)-1}{2}\right)x_j - \frac{x_j}{8}\left(\frac{f(D)}{\beta^H} + \frac{1-f(D)}{\beta^L}\right)$$

Differentiating with respect to D:

$$\frac{dEU_A^*(D)}{dD} = \frac{x_j}{2}m'(D) - \frac{x_j}{8}\left(\frac{1}{\beta_H} - \frac{1}{\beta^L}\right)f'(D)$$

Because $\frac{1}{\beta^H} - \frac{1}{\beta^L} < 0$, the first term is positive while the second is negative, meaning social distance has opposing effects on expected utility.

Taking the second derivative:

$$\frac{d^2EU_A^*}{dD^2} = \frac{x_j}{2}m''(D) - \frac{x_j}{8}\left(\frac{1}{\beta_H} - \frac{1}{\beta^L}\right)f''(D)$$

Because the decline in trust accelerates with distance (i.e. $f'(D) < 0$) while gains from the multiplier increase slowly,

$$\frac{d^2 EU_A^*}{dD^2} < 0$$

Thus expected utility is concave in social distance, implying that both very low and very high levels of social distance lead to lower welfare outcomes.

Reconnecting Young People

Joshua Frolish

Abstract

With loneliness increasingly recognised as an issue within our society, shown by the "A Connected Society" national strategy created to combat the issue, policymakers must understand the issue's causes to respond appropriately. This article aims to examine the causes of loneliness in young people and use the findings to suggest policy prescriptions to combat the issue. This article argues that mental health conditions are the most significant driver of loneliness in young people, and then recommends both tackling mental health directly through increasing capacity in the NHS for care, as well as tackling issues that exacerbate mental health issues and structural causes of loneliness. Policies to tackle these issues could include increasing the regulation of social media and offering targeted economic interventions, such as subsidised public transport, to ease the costs of socialising for young people. These actions will provide the holistic policy response necessary to effectively reconnect young people.

What is the Issue?

To begin, it is important to gain a greater understanding of the issue of loneliness in the UK and examine what existing studies have uncovered. While loneliness clearly impacts people of all ages, it is young people that appear to be the most affected by loneliness, "with 29% of those aged 16 to 24 feeling lonely at least some of the time compared with 15% of those aged 65 to 74" (NHS England, 2026). Therefore, this report focuses its policy towards young people to target the group most affected by the issue of loneliness.

Loneliness is now increasingly recognised as an important policy issue by successive governments, with the first strategy to combat the issue, "A Connected Society", published in 2018. The paper primarily recommends treating loneliness as a measurable public health issue, with better data collecting with standardised questions and measurements. This would support an improved policy response focused on connecting people to community organisations through the NHS, with support targeted at key transition stages in life where the risk of loneliness is greatest. The report also emphasised cross-government cooperation and efforts to reduce stigma (DDCMS, 2018). Following this report, the government has taken action to reduce levels of loneliness, with information campaigns such as 'Let's Talk Loneliness', as well as £5 million given to national organisations that were already tackling loneliness (DDCMS, 2022). It seems, then, that loneliness is established as an issue which ought to be, and already is, mitigated by public forces. This report will discuss how to further mitigate loneliness in young people.

What's Causing the Issue?

When examining the risk factors for loneliness in young people, it clearly appears that the most significant is having poor mental health. In 2023, only 5.2% of young people (those aged 17 - 22) without a mental disorder reported often or always feeling lonely, with this figure rising to 29.5% in those who likely suffered from one (NHS England, 2023). This data shows the significant impact that poor mental health has on loneliness, demonstrating the

importance of any policies which can improve young peoples' mental states. Other data shows that poor mental wellbeing can initiate and lengthen periods of chronic loneliness, with individuals who suffer from mental distress being 4.2 times as likely to report chronic loneliness than those without. Despite this, the phenomenon is bidirectional, with loneliness degrading people's mental health. This can be observed in the data as those who experienced prior chronic loneliness became 3.7 times more likely to report mental distress than those who weren't (DDCMS, 2022). Therefore, while policies directly targeting mental health will surely be essential, there is the importance of tackling the root causes of poor mental health to develop a holistic response. Therefore, we must also discern what the risk factors for poor mental health in young people may be.

Social media has been linked to poor mental health with "27% of children spending 3+ hours daily on social media show[ing] mental health symptoms, compared to 12% who spend none" (Mind, 2025). This data appears to support other research which has indicated that lonely individuals are more prone to engaging with social media in more problematic and passive ways (Emily B. O'Day, Richard G. Heimberg, 2021). Usage in this way often exacerbates poor mental health outcomes, particularly when individuals compare their lives to others on social media - which lonely people, who are unsatisfied with their social lives, may be likely to do. Thus, one way to protect the mental health of young children could be to better regulate usage of social media platforms. These regulations may help to alleviate a risk factor for loneliness, and would also limit some of the most problematic usage of social media. This would defend the mental health of lonely people which, alongside increasing their quality of life directly, may also help reduce the amount of time these individuals suffer from chronic loneliness.

Alongside this, however, many broader societal factors impact upon mental health. For example, lockdowns during the Covid-19 pandemic appeared to produce a dramatic spike in mental health disorders - with a 47.8% increase in 17-19 year olds having a probable mental disorder between 2021-22 (NHS England, 2022). Additionally, lower income levels are also a major risk factor for poor mental health with the ONS reporting that "Children in relative poverty had almost twice the odds of reporting feeling lonely" (ONS, 2018). Therefore, a holistic approach must tackle all of these issues to have a truly significant impact upon young people.

How Can We Solve the Issue?

Now that the causes of loneliness have been identified, it is possible to recommend policy solutions that actually tackle these root issues and will thus, hopefully, be more effective in tackling loneliness. It appears that the most important factor behind loneliness in young people is poor mental health. Therefore, a holistic approach to resolving the mental health crisis is necessary for reducing levels of loneliness in young people.

Firstly, the government ought to implement a direct strategy to improve the mental health of young people. When surveyed by 'Mind', "41% of young people surveyed said waiting times were the biggest barrier to support" (Mind, 2025). This shows a simple truth that there is a lack of capacity in mental health services for young people. Ultimately, improving access and capacity in these services will likely require a targeted increase in funding to reduce the backlog of waiting lists. This will allow young people to access the professional services they need to overcome or manage their mental health disorders. It is essential to note that this is seemingly not currently being prioritised in NHS budgets as the proportion of spending on mental health has fallen this year (NHS England, 2025) - suggesting that policymakers are currently underprioritising this issue.

NHS surveys have shown that rates of probable mental disorders amongst 17-19 year olds have risen from 10% in 2017 (NHS England, 2022) to 23.3% in 2023 (NHS England, 2023), an increase of more than 100% over the period. While improved diagnoses may account for part of this increase, it is clear that mental health issues have worsened over time which ought to warrant greater attention in order to return levels to their previous rate. Additionally, it must be said that the lack of spending on mental healthcare creates long-term costs with the societal cost of mental ill-health estimated at £300 billion in 2022 (Centre for Mental Health, 2024). Crucially, that is an almost £200 billion increase from 2014 when the figure stood at just £105 billion (NHS England, 2016). Therefore, in the long-term, investment in this care will ultimately save more money than is spent.

Following this increase in investment, the government should implement further policies to mitigate the root causes of poor mental health to prevent future issues and ensure improvements in mental health are sustained. One action to take would be regulation of social media. There have been proposals to ban social media platforms for under-16s and some

jurisdictions have even implemented this policy. However, this will likely be difficult to enforce and a more limited set of proposals may be more effective unless future data suggests the total bans did work effectively. More limited regulatory steps can be taken to curb the addictive impacts of social media - with less time spent on the apps potentially reducing their negative impacts. For example infinite scroll features could be banned, a measure EU preliminary research has already recommended for TikTok (European Commission, 2026). This would help to reduce addictive loops of 'doomscrolling' which keep many users glued to their screens. A reduction in passive social media usage, identified as problematic, would help to protect the mental health of lonely young people as well as reducing levels of addiction to social media.

Another action is to rebuild social centres in communities to heal some of the lack of socialisation caused by the pandemic lockdowns. Children have suffered from undersocialisation due to not being in school, while being simultaneously affected by a lack of youth services and early intervention programs. Real funding for these measures has declined by 73% from 2010-11 to 2023-24 (Mind, 2025). Combined, these changes amount to a serious threat to young people's mental health. To redress this, the defunding of local services for young people must be reversed at least partially.

This funding could be channeled through a combination of local councils, schools and local charitable organisations/NGOs. Councils could support community centres, leisure centres and sports clubs. Meanwhile, schools could use funding to provide their pupils with a variety of clubs and activities which could provide them easy opportunities to socialise with their peers during the week. Additionally, it is crucial to support the variety of NGOs that are vital to providing opportunities to socialise in local communities. For example, the Porch Project in Hadleigh - a small town in Suffolk - provides a safe space for young people to socialise outside of school hours. On a larger scale, organisations such The Scout Association provide social opportunities to children across the country. Increasing the funding available to these actors would increase the accessibility of social opportunities for young people and support the development of the communication skills and friendships that will help to reduce levels of loneliness down the line. The government has already funded national organisations targeting loneliness (DDCMS, 2022) but local councils could also be given funding to distribute to the variety of important projects in their communities that create and maintain the social networks necessary to reduce loneliness. Ultimately, support for community organisations

cannot be reduced as levels of probable mental disorders and loneliness rise if the government genuinely aims to improve the lives and mental states of young people.

A third response could be a series of economic measures to alleviate cost of living issues which impact upon young people's abilities to socialise. While broader measures to redress major economic issues, such as underemployment or inequality, are beyond the scope of this report, there are some novel economic measures being suggested that could alleviate some of the economic barriers to young people socialising. One example is the 'culture card' that Ireland may soon introduce (Fine Gael Press Office, 2024). The €100 card would be spent by 18 year-olds on live events, providing economically disadvantaged young people an opportunity both to experience culture, but also to socialise with their peers in events which may otherwise have been unaffordable. Another innovative proposal is that by the UK Green Party to make bus travel free for under-22s (BBC, 2026). This would assist young people, particularly in more rural areas, to reach any social events.

Despite their high price tags, these programs would do a lot to mitigate structural factors which contribute to loneliness. For example, many areas suffer from a lack of accessible third spaces (areas to socialise away from both the home and the workplace). This issue is particularly acute for young people who often lack the financial resources to access the third spaces which are available. These example proposals support both sides of the issue. They reduce the cost to access these spaces, directly making them more accessible to young people who would otherwise lack the economic means. Furthermore, though, this would increase the numbers of people utilising third spaces overall. This may increase the incentive to open and run such places, which may ultimately increase the general availability of third spaces. These effects would encourage greater socialisation for young people which ought to reduce levels of loneliness as it is easier for individuals to fulfill their demand for interaction with other people.

Conclusion

In conclusion, the most significant cause of loneliness in young people appears to be mental health issues, but the presence of broader structural issues and the possibility of a bidirectional effect make a holistic approach especially important. To remedy the issue, policymakers ought to pursue this holistic approach, not only tackling mental health directly

through increased NHS capacity, but also taking preventative measures by alleviating other structural drivers of poor mental health and loneliness. These efforts could include innovative policies to alleviate cost of living issues as well as targeted regulation of social media to curb its most harmful areas. Additionally, young people deserve local communities with thriving social networks they can participate in and therefore the large cuts to youth services and early intervention programs ought to be reversed, at least in part. These measures are likely to succeed as they target one of the primary causes of loneliness directly, while also attacking structural issues, leaving our young people with not just the support they need, but also a wider society which is easier to socialise in generally. The actions will require significant investment however, requiring the end to underprioritisation, reversals of defunding as well as greater innovation with regards to policy measures. Despite this, the costs of loneliness and poor mental health, both to social wellbeing and the economy are clear, and this spending ought to be viewed as a crucial investment into the future of our national wellbeing which will pay dividends for years to come.

Bibliography

Centre for Mental Health (2024). *The economic and social costs of mental ill health*. [online] Centre for Mental Health. Available at: https://www.centreformentalhealth.org.uk/publications/the-economic-and-social-costs-of-mental-ill-health/ [Accessed 2 May 2026].

Digital.nhs.uk. (2026). *Loneliness and Wellbeing*. [online] Available at: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2024/loneliness-and-wellbeing [Accessed 2 May 2026].

European Commission - European Commission. (2026). *Commission preliminarily finds TikTok's addictive design in breach of the Digital Services Act*. [online] Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_26_312 [Accesssed 2 May 2026].

For, D. (2022). *Investigating factors associated with loneliness in adults in England*. [online] GOV.UK. Available at: https://www.gov.uk/government/publications/factors-associated-with-loneliness-in-adults-in-england/investigating-factors-associated-with-loneliness-in-adults-in-england#how-does-loneliness-relate-to-mental-wellbeing [Accessed 2 May 2026].

FUSIO and Fine (2024). *Fine Gael sets stage for young people to experience theatre, arts and culture – Sammon - Fine Gael*. [online] Fine Gael. Available at: https://www.finegael.ie/fine-gael-sets-stage-for-young-people-to-experience-theatre-arts-and-culture-sammon/ [Accessed 2 May 2026].

GOV.UK (2022). *Government's Work on Tackling Loneliness*. [online] GOV.UK. Available at: https://www.gov.uk/guidance/governments-work-on-tackling-loneliness [Accessed 2 May 2026].

HM Government (2018). *A Connected Society: A Strategy for Tackling Loneliness - Laying the Foundations for Change*. [online] Available at: https://assets.publishing.service.gov.uk/media/5fb66cf98fa8f54aafb3c333/6.4882_DCMS_Loneliness_Strategy_web_Update_V2.pdf [Accessed 2 May 2026].

Mind.org.uk. (2019). *Care Before Crisis: our report on young people's mental health*. [online] Available at: https://www.mind.org.uk/news-campaigns/campaigns/children-and-young-peoples-mental-health/care-before-crisis-our-report-on-young-peoples-mental-health/ [Accessed 2 May 2026].

Morton, B. (2026). Green Party calls for free bus passes for under-22s. *BBC News*. [online] 17 Jan. Available at: https://www.bbc.co.uk/news/articles/cpv709w9j1xo [Accessed 2 May 2026].

NHS England (2024). *NHS England» Mental Health Five Year Forward View Dashboard*. [online] England.nhs.uk. Available at: https://www.england.nhs.uk/mental-health/taskforce/imp/mh-dashboard/ [Accessed 2 May 2026].

NHS Digital (2022). *Mental Health of Children and Young People in England 2022 - Wave 3 Follow up to the 2017 Survey*. [online] NHS Digital. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2022-follow-up-to-the-2017-survey [Accessed 2 May 2026].

NHS Digital (2023). *Mental health of children and young people in england, 2023 - wave 4 follow up to the 2017 survey*. [online] NHS Digital. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2023-wave-4-follow-up [Accessed 2 May 2026].

NHS (2023). *Part 2: Sleep, loneliness, Activities and Health Behaviours*. [online] NHS Digital. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2023-wave-4-follow-up/part-2-sleep-loneliness-activities-and-health-behaviours [Accessed 2 May 2026].

O'Day, E.B. and Heimberg, R.G. (2021). Social media use, social anxiety, and loneliness: A systematic review. *Computers in Human Behavior Reports*, [online] 3(100070). doi:https://doi.org/10.1016/j.chbr.2021.100070.

Ons.gov.uk. (n.d.). *Exploring loneliness in children, Great Britain - Office for National Statistics*. [online] Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/exploringlonelinessinchildrengreatbritain/2018 [Accessed 2 May 2026].

The 'Women and Children first myth': the growing harms of misogyny and deepfakes against women and girls in the UK

Mia Sadler

Abstract

This article examines recent steps taken by the Labour government to address creation of 'purported intimate images' and deepfakes. It will outline wider political discourse around this issue alongside previous and proposed legislation to evaluate whether these steps act as an effective tool in protecting women and girls against non-consensual creation of deepfakes.

Introduction

Misogyny and gender bias existing online is not a new phenomenon, yet it has been intensified with increased use and reliance on AI. Whilst its purposes are multifaceted, creation of deepfakes is a particularly harmful aspect, and the gendered nature of it cannot be understated. It is known that women are proportionately more affected than men, but the extent of it remains under researched, and this reflects in policy decisions. This article will place the amendments made to the 2003 Sexual Offences Act in the 2025 Data (Use and Access) Act of 2025 in a wider social and political context, assessing how effective they are in addressing current AI use and misogyny, and further action they should take.

The use, commercialisation and reliance on AI has been gradually increasing, yet deepfakes have significantly increased annually over the last five years. For example, in the UK, 500,000 deepfakes were shared in 2023, and the government projected that 8 million were shared in 2025 (Gov.UK, 2025). The overall optimism towards integration of AI technology in wider society has masked issues that arise with these figures, and shows how, in just two

years, the rate of deepfake creation and distribution has become un-manageable. A 2026 analysis of output from X's AI chatbot, Grok, found that out of 20,000 generated images, 12,995 were sexualised images of adults and children (CCDH, 2026). As of January 2026, there are an estimated 3,002,712 images of this nature existing on the platform (CCDH, 2026). The relationship of this to gender bias and misogyny has only recently started to be addressed, but more work needs to be done. The Alan Turing Institute is one organisation that is doing research on the danger of deepfakes, finding that whilst 15% out of 1403 participants have been exposed to deepfakes, 90.4% of them reported being concerned about the presence of them (Sippy et al, 2024). Within this, women were found to be more concerned than men, and it is this perception and public conscience that reflects wider issues at play (Sippy et al, 2024).

Re-centring deepfakes

The prompt increase of deepfakes reveals the need for governments to centre women and children in the issue. 99% of people targeted in deepfake pornography are women, and this genre makes up 98% of all deepfake videos online (UN Women, 2025). This shows that women and female identities are consistently exploited online, and that these are increasing on an alarming scale. Combining domestic data from the UK with wider findings does highlight the main issue with combatting deepfakes and their harms. Because they are so prominent on social media platforms like X (outlined above), it is difficult to assess in isolation what the UK government could do to solve the wider issues. However, the above data, paired with the findings about societal concerns about deepfakes, points to solutions that the UK government may implement leading on from the Data (Use and Access) Act, which will later be proposed. The need for the government to take consistent steps regarding deepfakes links to the existence of misogyny and the normalisation of gender violence on the internet. Misogyny and hatred of women online promoted by male influencers falls outside of legislation and legal jurisdiction currently, and there is not enough agency on governments to monitor this (Gibson, 2025). This is an issue, because it allows for the nature of deepfakes to continue to incite and normalise gender-based violence, and calls for deeper, meaningful solutions. Additionally, the online practice gives perpetrators the illusion that their actions come without consequences: if they generate an image, who will know, and how?

Politicians as a primary target

The study done by the Alan Turing Institute also found that 34.1% of deepfakes that people were exposed to were of politicians (Sippy et al, 2024). Whilst they were secondary to celebrities, it adds a nuance to the gendered argument and provides an insight for governmentally approaching the issue. For example, Alexandria Ocasio-Cortez (AOC) has been an advocate for combatting deepfake pornography, and acts as a symbol of the gendered issue at discussion. In 2024, she introduced the DEFIANCE Act, which allows for victims to take civil action against people who "produced, distributed, or received digital forgeries" (House.gov, 2024). Her experience as a victim of politically motivated deepfake pornography is indicative of a wider issue, but one that can evidentially be addressed by governments. Whilst this bill acted as a vital step for the US government in approaching the issue, the structure of US policy and agency of federal states limits the effectiveness. As other bills, the DEFIANCE Act has no federal protections and is unlikely to be revisited during the presidency of Trump, who frequently uses deepfakes to push false truths or political narratives. The above example demonstrates both a difference and comparison in approach to this issue. The use of deepfakes to create alternate political narratives is not uncommon, but including the discussion about female politicians is important for wider discussion. The DEFIANCE Act also acts as a point of comparison of approach, and a benchmark for the current Labour government.

Existing UK efforts

Compiling the above issues, the UK government has taken effective steps in limiting the harm of deepfakes. Despite not explicitly centring women in policy unlike the DEFIANCE Act, the Data (Use and Access) Act, Section 138, acts as a cornerstone. This amendment to the 2003 Sexual Offences Act in collaboration with other institutions aims to outline a clearer definition for deepfakes, and for consent. It also addresses deepfake pornography and intimate deepfakes, outlining specific actions such as requesting them as offences (Gov.UK, 2025). Moreover, there have been further pushes to address this issue through work done by The Accelerated Capability Environment (ACE). Scientific expertise has been used to begin processes of detection, creating dataset models and focussing these onto child sexual abuse material (ACE, 2025). They have also collaborated with tech companies for practical support, including Amazon and Microsoft. This is effective, in terms of company awareness.

As of now, the government has pushed forward more plans to use these detection tools and put more pressure on platforms to prevent deepfakes from being generated. Specifically, these

include ‘nudification’ tools provided to platforms (Home Office, 2026). These recent pledges have begun to focus on women and girls, with contributions from female MPs, but limits effectiveness for the wider public. There remains a gap between government action and public understanding. The general public, notably women, who still face fears about non-consensual deepfakes still have little understanding about how to bypass or alleviate these. They show attempts to keep up with advancements in AI technology, but these must remain consistent. Section 138 also sidelines harms faced by children and fails to outline clear and effective policing techniques.

Potential solutions

One of the best solutions for tackling deepfake harm is increased media literacy through established programmes. Media literacy is multifaceted, but focusing on the key aspects of critical thinking and technical understanding would allow for awareness and a sense of social responsibility (LIS Education Network, 2024). Access to these skills as part of a school curriculum or established courses would aid understanding of deepfakes and instil a sense of self-detection. Structuring courses by the key elements needed for media literacy could allow for optimal transmission of skills: critical thinking, analysis, questioning headline perspectives, production skills, and other ethical components. Knowing the harms of deepfakes and legislative consequences is also important for deterring perpetrators from creating them.

Equally, it is important for governments to integrate an intersectional approach when it comes to gendered issues. Acknowledging that non-white and non-heterosexual women suffer more than white, heterosexual women is significant and should be applied to the issue of deepfakes. As evidenced with the experiences of U.S. Representative Alexandria Ocasio-Cortez, normalised violence spread through deepfakes is an international concern, and impacts marginalised women. The UN Women’s AI school acts as a solution that, if widely implemented, can be pivotal in preventing harmful deepfakes. It is currently offered to women’s rights organisations, where they learn how to use AI tools responsibly and ethically and as a way to prevent further violence against women through AI (UN Women, 2025). This, combined with specific focus on media literacy, can be just as effective as policy change. This is what government attention should focus on to maximise how far policy can reach.

Conclusion

To conclude, these actions taken by the UK government show a clear move to address deepfakes and violence against women. Steps already taken, including utilisation of scientific expertise and detection tools, as well as pressuring platforms more, will monitor the creation and distribution of deepfakes. Action taken in the US shows how it is a universal and global issue that is starting to be taken seriously by major political powers. Calls to action deliver confidence that the current Labour government sees the issues and potential harms of deepfakes, and how women are impacted. However, given the complexity of digitality and the status of AI today, policies fall short of what is tangibly needed for the societal body. Alternative solutions must be considered to combine with current ones. By adopting educational courses proposed by the UN and by incorporating media literacy, people can use their agency to recognise the unknown damage of deepfakes and false news. AI technology is constantly evolving and adapting to the needs of a growing society, yet media literacy is constantly required. By giving this to people, they no longer must rely on policies that cannot hope to keep up. Efforts change, yet this can be the constant thread that prevents violence against women and girls in our technical age.

Bibliography

Accelerated Capability Environment, (2025). *Innovating to detect deepfakes and protect the public.* [Online] Available at: https://www.gov.uk/government/case-studies/innovating-to-detect-deepfakes-and-protect-the-public [Accessed 01 March 2026].

Ashikuzzaman, M., (2024.) *The Complete Guide to Media Literacy: Why It Matters Today.* [Online] Available at: https://www.lisedunetwork.com/the-complete-guide-to-media-literacy-why-it-matters-today/ [Accessed 02 March 2026].

Center for Countering Digital Hate, (2026.) *Grok floods X with sexualized images of women and children.* [Online] Available at: https://counterhate.com/research/grok-floods-x-with-sexualized-images/ [Accessed 04 March 2026].

Gibson, K., (2025.) *Gender bias, AI, and deepfakes are promoting misogyny online.* [Online] Available at: https://blogs.lse.ac.uk/wps/2025/01/09/gender-bias-ai-and-deepfakes-are-promoting-misogyny-online/ [Accessed 01 March 2026].

Gov.UK, (2026.) *Government leads global fight against deepfake threats.* [Online] Available at: https://www.gov.uk/government/news/government-leads-global-fight-against-deepfake-threats [Accessed 02 March 2026]

Library of Congress , (2024.) *Text - S.3696 - 118th Congress (2023-2024): DEFIANCE Act of 2024."*. [Online]
Available at: https://www.congress.gov/bill/118th-congress/senate-bill/3696/text [Accessed 01 March 2026].

Sippy, T., Enock, F. E., Bright, J. & Margetts, H. Z., 2024. Behind the Deepfake: 8% Create; 90% Concerned. *Public Policy Programme ,* pp. 1-9.

UN Women, (2025.) *AI-powered online abuse: How AI is amplifying violence against women and what can stop it.* [Online]
Available at:

https://www.unwomen.org/en/articles/faqs/ai-powered-online-abuse-how-ai-is-amplifying-violence-against-women-and-what-can-stop-it [Accessed 01 March 2026].

Who Really Decides... AI or Humans?

AI, Operational Decision-Making, and Silent Disconnection in Organisations.

Joshua Taylor

Abstract

Artificial intelligence is increasingly embedded in organisational decision-making, shaping processes such as labour scheduling, performance monitoring, and financial risk assessment. While these systems enhance efficiency and analytical capacity, their growing autonomy and adaptivity introduce new governance challenges. This article examines how AI reshapes organisational authority and explores the mechanisms required to maintain institutional legitimacy as algorithmic systems become embedded in operational processes. Drawing on empirical research and institutional reports, the article argues that improved AI capability can paradoxically centralise managerial oversight rather than eliminate it, while algorithmic management contributes to declining employee trust and what this paper terms silent disconnection. Evidence from financial services further demonstrates how reliability concerns and regulatory fragmentation constrain high-risk deployment. The article concludes that transparency alone cannot restore legitimacy; effective governance requires contestability, human oversight, and coordinated regulatory frameworks that align technological capability with institutional accountability.

Introduction: Adaptivity, Autonomy, and the New Opacity of Authority

Who determines shift patterns, flags employee performance risks, or authorises a loan approval? Increasingly, these decisions are not made through visible managerial deliberation but through algorithmic systems embedded within organisational infrastructure.

Artificial intelligence now resides within the operational core of organisations. It schedules labour, optimises supply chains, predicts turnover, allocates credit risk, and monitors compliance. In structured environments, AI systems outperform humans in speed, scale, and pattern recognition. Efficiency improves. Costs decline. Variability narrows.

Alongside these efficiency gains, a deeper organisational transformation is emerging.

The modern AI system is defined by two technical characteristics: adaptivity and autonomy. Adaptivity refers to the ability of machine-learning systems to update their outputs through probabilistic inference and iterative training cycles, often producing decision logic that is difficult for human operators to interpret. Autonomy means that systems execute decisions without continuous human intervention. When adaptivity makes logic inexplicable and autonomy makes responsibility difficult to assign, authority becomes embedded in technical architecture rather than human judgement.

The result is not an immediate breakdown. The deeper risk is more subtle. It is a silent disconnection. This shift matters because algorithmic decisions increasingly shape

employment, financial access, and organisational governance across advanced economies. Individuals comply with algorithmic decisions yet disengage from the processes that produce them. Responsibility becomes diffuse. Control becomes opaque.

Extracting contemporary research and institutional reports, this article argues that AI restructures organisational hierarchies, centralises interpretive authority, narrows discretion, and creates regulatory friction. If governance does not evolve alongside capability, high performance systems will lack durable legitimacy. To understand this shift, we begin with structure.

This article therefore asks: **how does artificial intelligence reshape decision-making authority within modern organisations, and what governance mechanisms are required to preserve organisational legitimacy as algorithmic systems become embedded in operational processes?**

I. Structural Paradox: Reliability and the Narrowing of Control

Supportive accounts argue that AI broadens access to expertise by reducing the cognitive demands traditionally required to perform complex tasks. In theory, advanced systems enable less experienced workers to perform complex tasks through augmentation. This view implies flatter hierarchies and wider spans of control. Emerging evidence complicates that assumption.

Xu et al. (2025) identify a structural paradox that directly challenges traditional expectations of technological flattening. As generative AI reliability improves, organisations do not necessarily decentralise authority. Instead, managers must act as interpretive gatekeepers who validate outputs and mitigate hallucination risks. The more capable the system becomes, the more responsibility consolidates around those authorised to verify it. Rather than eliminating oversight, improved system capability often intensifies supervisory control.Enhanced reliability can therefore narrow the span of control because managerial validation becomes central to risk containment (Xu et al., 2025, pp. 14 -15). In this sense, AI reverses the flattening logic associated with earlier information technologies. Knowledge may be distributed, but accountability is re-centralised.

This produces what may be termed the deskilling paradox. Routine knowledge requirements decline, but interpretive authority consolidates upward. Frontline workers rely on AI outputs. Managers verify, escalate, and assume liability. Hierarchies do not disappear. They reconfigure around algorithmic oversight. Authority becomes less visible yet more concentrated.

This structural centralisation shapes how power is experienced in daily work.

Traditional Decision Structure	**AI-mediated Decision Structure**
Human judgement central	Algorithmic output central
Manager decides	Manager verifies
Worker discretion present	Worker executes AI outputs
Authority visible	Authority embedded in systems
Accountability individualised	Accountability diffused

Source: Author framework based on Xu et al. (2025) and Uddin (2026).

II. Invisible Management and the Erosion of Trust

Algorithmic management systems translate structural centralisation into operational practice. Tasks are allocated through automated scheduling. Performance is monitored through real time analytics. Discipline may be triggered by threshold breaches embedded in code.

Uddin (2026) provides empirical evidence of the consequences. Using survey and interview data, the study identifies a strong negative relationship between intensive digital surveillance and employee trust (Uddin, 2026, pp. 45 - 47). Workers subjected to algorithmic monitoring report diminished discretion and heightened anxiety. Uddin (2026) further demonstrates that this relationship is not marginal but structurally significant. Intensive surveillance correlates with measurable declines in perceived fairness and autonomy. Employees report compliance driven by system inevitability rather than managerial trust. When control is exercised without visible human reasoning, authority appears unchallengeable. The result is not resistance but psychological withdrawal.

This condition is described as invisible management. Authority is no longer embodied solely in a human supervisor but embedded within dashboards, metrics, and algorithmic systems. Individuals comply because deviation is measurable and penalised. However, trust declines when decision logic remains opaque and recourse is limited. Silent disconnection emerges here. Workers follow instructions generated by systems they neither designed nor fully understood. Compliance replaces commitment. Efficiency is preserved while legitimacy weakens. The question becomes how legitimacy can be restored when decision authority shifts to adaptive systems.

III. The Appeal Dividend and the Limits of Transparency

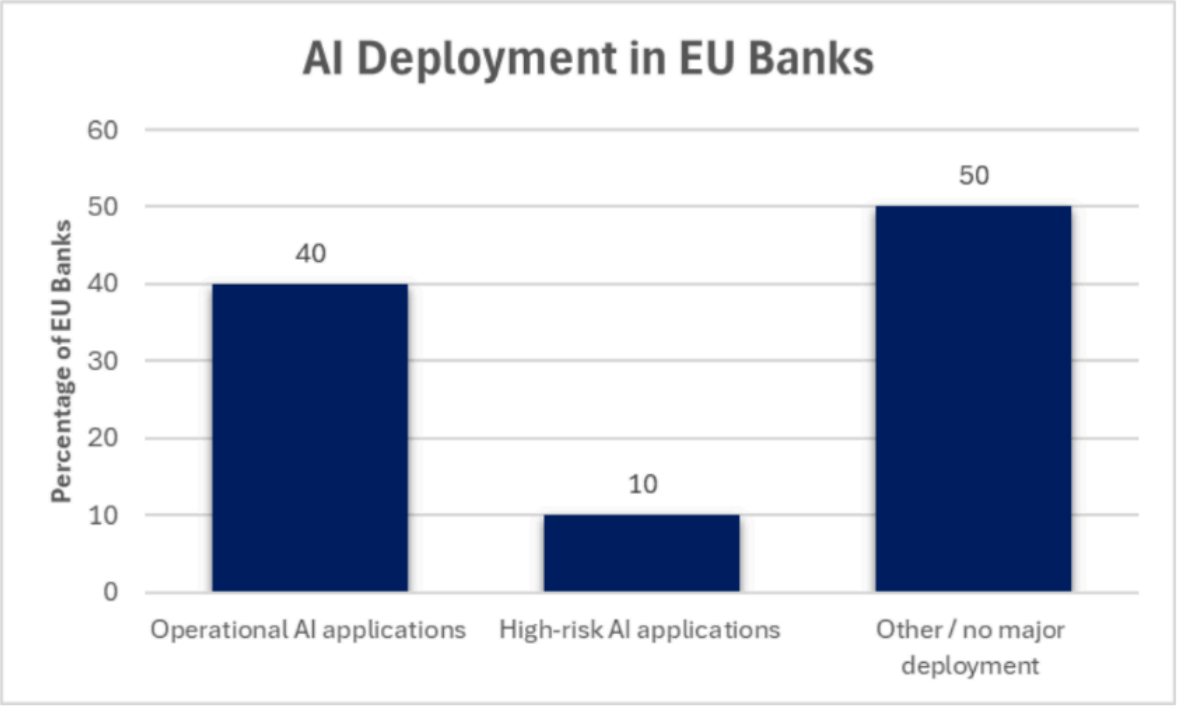
A common organisational response to algorithmic opacity is transparency. Firms disclose that automated systems are used. They provide explanatory statements or model summaries. Empirical evidence suggests this approach is insufficient. Martin and Waldman (2022) demonstrate that standard transparency notices do not increase perceived legitimacy.Martin and Waldman (2022) demonstrate that standard transparency notices do not increase perceived legitimacy. In some cases, disclosure that a decision was algorithmically generated produces a legitimacy penalty (Martin and Waldman, 2022, pp. 346
- 347). Individuals perceive algorithmic decisions as reductionist when procedural safeguards are weak.

The only governance mechanism that consistently produces a legitimacy dividend is a formal human appeal process (Martin and Waldman, 2022, pp. 346 - 347). When individuals have access to meaningful human review, even adverse outcomes are perceived as more Legitimate. Transparency alone does not generate trust. Contestability does. This finding fundamentally reframes AI governance debates. Many institutions prioritise disclosure as a compliance mechanism. Yet empirical evidence shows that information without agency fails to restore legitimacy. The decisive variable is not awareness but recourse. Where appeal is absent, algorithmic authority appears final. Where appeal is present, institutional legitimacy stabilises even under adverse outcomes. This finding has direct implications for sectors where AI decisions materially affect economic opportunity.

IV. Sectoral Evidence: Reliability Friction in Banking

The European Banking Authority (2024) reports that approximately 40 percent of EU banks deploy AI in operational contexts, including profiling and customer service optimisation. However, only around 10 percent apply AI in high risk functions such as credit scoring and anti money laundering enforcement.

Figure 1. AI deployment across EU banks



Source: European Banking Authority (2024); author visualisation based on cited data.

This disparity reflects what can be described as reliability friction. Banks express concern about hallucinations, explainability constraints, and consumer protection risks (European Banking Authority, 2024, pp. 89 - 91). Adaptive systems can generate outputs that are statistically coherent yet contextually problematic.

Consequently, institutions reconstruct decision authority around risk-based governance guardrails. Human oversight intensifies in high stakes decisions. AI assists but does not replace final judgement. However, embedding such oversight within overlapping regulatory frameworks introduces additional burdens. These adoption patterns illustrate that the central challenge of AI governance is not technological capability but institutional confidence in decision accountability.

V. Regulatory Fragmentation and the Chilling Effect

AI deployment in Europe must comply with multiple regulatory regimes. Graux et al. (2025) identify significant regulatory complexity arising from the interplay between the AI Act and the General Data Protection Regulation.

Deployers may be required to conduct both Data Protection Impact Assessments and Fundamental Rights Impact Assessments for the same system. Logging, transparency, and documentation obligations overlap, creating redundant requirements (Graux et al., 2025, pp. 1043, 1119). These duplications increase compliance costs while simultaneously generating administrative uncertainty for deployers. Industry reaction illustrates the consequence.

A coalition of firms issued the so-called Stop the Clock protest, arguing that unclear risk categorisation and overlapping obligations generate a chilling effect on innovation. This fragmentation produces a broader form of silent disconnection at the institutional level. Innovators disengage from experimentation not because risk is unacceptable, but because compliance ambiguity is unpredictable. Overlapping obligations for logging, transparency, and documentation creates redundant governance layers that increase legal uncertainty. As a result, authority shifts from operational experimentation toward defensive legal interpretation. The chilling effect is therefore structural rather than rhetorical. Innovation slows not due to technological limitation, but due to institutional incoherence. The challenge is therefore to coordinate governance rather than dilute it.

VI. Institutional Co-production and a Coordinated Governance Model

An alternative governance framework is provided by Hao et al. (2025). Their study of AI deployment in aviation identifies epistemic asymmetry between statistical models and contextual reasoning. AI systems handle optimisation and structured logic but lack semantic specificity during operational volatility (Hao et al., 2025, Article 101152).

Policy/Model	Pros	Cons	Risk Mitigation
Current Fragmented Governance	High operational speed, reduced variability, and immediate cost declines in structured tasks.	Opaque logic, "invisible management," redundant compliance burdens, and a chilling effect on innovation.	Reliance on defensive legal interpretation and standard transparency notices (often ineffective).
Institutional Co-Production	Restores institutional legitimacy, ensures contextual reasoning, and accelerates market access.	Intensified supervisory requirements for managers acting as interpretive gatekeepers.	Mandatory human appeal processes and coordinated compliance architectures to reduce redundancy.

The authors propose institutional co-production. Under this model, AI generates preliminary outputs while humans provide contextual interpretation, ethical framing, and adaptive judgement. Authority is therefore shared but functionally differentiated. Machines optimise. Humans interpret.

Governance must therefore embed co-production rather than assume substitution. Institutional co-production should be treated as a design principle rather than an optional safeguard. Governance frameworks must explicitly allocate interpretive authority to human actors in contexts of volatility, ambiguity, or ethical sensitivity. AI systems should optimise structured reasoning, while humans retain responsibility for contextual judgement and normative evaluation. Embedding this division prevents authority from collapsing entirely into statistical abstraction. Policy should incorporate three coordinated elements.

First, mandatory human appeal mechanisms for consequential decisions, ensuring contestability as demonstrated by Martin and Waldman (2022).

Second, regulatory sandboxes to reduce compliance uncertainty. The UK Financial Conduct Authority digital sandbox accelerated speed to market by approximately 40 percent for participating firms (UK Government, 2023, p. 173). Sandboxing allows SMEs to innovate under supervision without prohibitive risk.

Third, a coordinated compliance architecture that aligns impact assessments and documentation requirements across regimes, reducing redundancy while preserving safeguards. Such measures preserve both efficiency and accountability.

Conclusion

AI has become embedded in the operational heart of organisations. These policy mechanisms illustrate that the challenge of AI governance is not whether humans or machines decide, but how decision authority is structured between them. It enhances forecasting, accelerates optimisation, and expands analytical capacity beyond human limits. Yet its defining characteristics, adaptivity and autonomy, render decision logic opaque and responsibility diffuses.

Empirical evidence demonstrates that improved AI reliability can narrow spans of control through intensified managerial gatekeeping (Xu et al., 2025). Intensive surveillance correlates with declining employee trust and invisible management (Uddin, 2026).

Transparency notices fail to restore legitimacy, whereas formal human appeal produces a legitimacy dividend (Martin and Waldman, 2022). Overlapping regulatory obligations create redundant burdens and a chilling effect on innovation (European Parliament, 2025). Sectoral adoption patterns in banking confirm that reliability friction constrains high risk deployment (European Banking Authority, 2024).

Silent disconnection is not an inevitable consequence of artificial intelligence. Rather, it reflects a structural misalignment between rapid technological capability and the slower development of organisational governance frameworks. Elghomri et al. (2025) demonstrate that approximately 66 percent of AI research focuses on theoretical or technical advancement, while only 23 percent addresses practical implementation and governance integration within organisations (pp. 281 - 283). This imbalance highlights a persistent gap between innovation and institutional readiness. As technological capability accelerates, governance mechanisms often lag behind, leaving organisations with powerful systems but insufficient structures to oversee their deployment responsibly.

The defining challenge of the automated age is therefore not computational power but

institutional alignment. If adaptivity renders logic opaque and autonomy diffuses responsibility, governance must restore interpretability and assignability. Contestability, oversight, and regulatory coherence are not administrative details. They are structural safeguards for legitimacy.

The question is no longer whether AI decides. It already shapes consequential outcomes across labour, finance, and supply chains. The decisive issue is whether organisations design systems in which human authority remains meaningful and accountable. Efficiency without legitimacy is unstable. Authority without recourse is fragile. Bridging the gap between capability and readiness is the central institutional task of AI governance.

Bibliography

Department for Science, Innovation & Technology and Office for Artificial Intelligence (2023)
A pro-innovation approach to AI regulation. Command Paper 815. London: HM Government.
Available at:
https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach (Accessed: 13 March 2026).

Elghomri, B., Messaoudi, F. and Touti, N. (2025) 'The role of AI in driving accountability and
transparency in global supply chains: The fragmented bridge between research and practice', Operations and Supply Chain Management: An International Journal, 18(2), pp. 275–287. doi: 10.31387/oscm0610472.

European Banking Authority (2024) 'Special topic – Artificial intelligence', in Risk Assessment Report of the European Banking Authority – November 2024. Paris: EBA, pp. 86–103. Available at:
https://www.eba.europa.eu/sites/default/files/2024-11/f03ee0c1-7258-4391-8bf1-578924956049/EBA%20Risk%20Assessment%20Report%20-%20Autumn%202024.pdf (Accessed: 13 March 2026).

Graux, H., Garstka, K., Murali, N., Cave, J. and Botterman, M. (2025) Interplay between the AI Act and the EU digital legislative framework. PE 778.575. Brussels: European Parliament Policy Department for Transformation, Innovation and Health. Available at:
https://www.europarl.europa.eu/supporting-analyses (Accessed: 13 March 2026).

Hao, X., Demir, E. and Eyers, D. (2025) 'Beyond human-in-the-loop: Sensemaking between artificial intelligence and human intelligence collaboration', Sustainable Futures, 10, 101152. doi: 10.1016/j.sftr.2025.101152.

Martin, K. and Waldman, A. (2022) 'Are algorithmic decisions legitimate? The effect of process and outcomes on perceptions of legitimacy of AI decisions', Journal of Business Ethics, 18(2), pp. 346–377. doi: 10.1007/s10551-021-05032-7.

Uddin, A. F. M. A. (2026) 'Trust under the algorithm: Employee perceptions of control, fairness, and autonomy in algorithmic management', Saudi Journal of Business and Management Studies, 11(2), pp. 40–52. doi: 10.36348/sjbms.2026.v11i02.003.

Xu, F., Hou, J., Chen, W. and Xie, K. (2025) ‘Generative AI and organizational structure in the knowledge economy’, arXiv [Preprint]. Available at: https://arxiv.org/abs/2506.00532 (Accessed: 13 March 2026).