AI-Assisted Political Persuasion: Evidence from 75 Online Campaign Conversations in the 2026 UK Local Election

Moses H. Siregar and John Bryden

June 2026

CAMPAIGN LAB

Executive Summary

Moses H. Siregar and John Bryden from CampaignLab developed and deployed an AI chatbot through Facebook Messenger as a primary campaign communication tool for a local councillor candidate during the 2026 UK local elections. The project examined whether generative AI could support political engagement and persuasion in a real electoral setting, and whether combining AI-led conversations with direct candidate intervention produced additional value.

The chatbot was built using a Python FastAPI backend, Facebook Messenger webhooks, Google Firestore for conversation and prompt storage, and an LLM accessed via the OpenRouter API. It was designed to hold short, conversational Messenger-style exchanges in British English, disclose its AI status, prompt residents to ask local questions, and route conversations to human follow-up when needed. Candidate-specific information was gathered through surveys, Facebook and website scraping, and local achievement data, allowing the chatbot to represent the candidate's background, record, and priorities.

The chatbot was promoted through organic Facebook posts, community group activity, and Facebook Ads. Ads framed the chatbot as a way for residents to ask questions, raise local concerns, review the candidate's record, and test a new form of accessible campaign communication. Three consent approaches were tested over the rollout: explicit chatbot-based consent, Facebook ad-based consent with a "No, thanks" option, and implicit consent through an opening disclosure message.

The analysis is based on 75 naturally occurring chatbot conversations. Persuasion was measured as an observable conversation-level outcome rather than assumed private attitude change. The main persuasion score ranged from -3 to +4 and captured visible indicators such as engagement, cooperation, trust, mobilisation, resistance reduction, relational warmth, and explicit support or rejection. Additional outcomes included mobilisation, trust and legitimacy, relational affinity, and resistance to change.

The findings suggest that AI's persuasive value depends less on the mere presence of a chatbot and more on the quality of the interaction. Across models, AI quality was the strongest and most consistent predictor of persuasion-related outcomes. Conversations were more persuasive when the chatbot responded empathetically, grounded answers in local issues, linked concerns to the candidate, prompted useful action, and helped discover the user's underlying issue.

Candidate intervention also appeared to add value. The candidate intervened in 19 of the 75 conversations, usually early in the exchange. Candidate-intervened conversations had higher descriptive persuasion, mobilisation, trust, and relational affinity scores than AI-only conversations. In the main hybrid model, candidate intervention remained positively associated with persuasion after adjusting for AI quality, consent phase, initial resistance, service requests, ideological dissent,

and AI scepticism. Selection-adjusted and inverse-probability weighted models produced similar results, although the study remains associational rather than causal.

The chatbot also functioned as a triage and routing mechanism. Candidate intervention was especially likely when users raised practical service requests, indicating that the system helped identify conversations requiring human authority, casework, reassurance, or accountability. Human follow-up appeared particularly valuable when users were sceptical, resistant, or uncomfortable engaging with automation.

One of the most important campaign insights came from the "No, thanks" consent option. Although initially appearing to signal disengagement, these refusals became useful indicators of hesitation or distrust. When the candidate personally followed up with users who had declined chatbot engagement, some conversations became more meaningful than the automated exchanges themselves. This suggests that refusal can operate as a reverse campaign funnel, identifying residents who may require direct human contact rather than automated persuasion.

The evidence points to an additive rather than synergistic hybrid model. AI quality and candidate intervention each contributed independently to persuasion-related outcomes, but there was no strong evidence that candidate intervention became more effective when AI quality was higher. The chatbot provided scale, accessibility, issue discovery, and triage, while the candidate provided legitimacy, accountability, local credibility, and practical responsiveness.

The study has important limitations. The sample size is small, candidate intervention occurred in only 19 conversations, and intervention was not randomly assigned. Persuasion was coded from conversation text rather than verified behavioural outcomes such as voting. Some variables were AI-assisted and require further manual validation. As a result, the findings should be interpreted as exploratory and associational.

Overall, the research supports a staged model of AI-assisted political persuasion. The chatbot acted as a persuasive gateway by opening conversations, sustaining engagement, identifying resident concerns, and creating opportunities for targeted human follow-up. Candidate intervention then added authenticity and trust where automation alone was insufficient. The strongest conclusion is not that AI chatbots directly persuade voters on their own, but that high-quality AI engagement can strengthen campaign communication when embedded within a hybrid system that routes the right conversations to human candidates at the right moment.

Background and Strategic Objectives

Given the growing relevance and accessibility of generative AI in political campaigns worldwide, many scholars and politicians have begun to explore the incorporation of such a tool within social media messenger features. In CampaignLab, we developed an innovative AI Chatbot embedded in the Facebook Messenger of a local councillor candidate, serving as their primary political campaign tool during the 2026 UK local election. By doing so, we aim to achieve the following strategic objectives:

1. To evaluate whether an AI chatbot can facilitate political persuasion and engagement in a real election campaign
2. To examine whether combining AI-mediated conversation with direct candidate intervention produces additional persuasive value
3. To understand the mechanisms through which AI contributes to campaign communication, including issue discovery, voter engagement, triage of constituent concerns, and routing users to human follow-up when necessary

Data and Methods

Building the AI Chatbot

We built an AI chatbot as a Facebook Messenger-based conversational tool connected to an LLM through a Python backend. Incoming messages from residents were received through a webhook, processed by the backend, and sent to the model together with a customised system prompt and recent conversation history. Conversation history and page-specific information were stored in Google Firestore.

Collaborating with a Local Councillor Candidate

We provided the AI Chatbot service to a local councillor candidate for the 2026 UK Local Election, who was willing to make the chatbot a primary means of campaigning. The candidate filled out a survey that gives us all the necessary information about the candidate's background, track record, and vision and mission. We also scraped data using Apify from the candidate's Facebook page and local party website, as well as used Wardwatch to compile a local councillor achievement report. All of this information is then input into the prompt field, so that the chatbot can maximise its representation of the local candidate.

Promoting the AI Chatbot

The AI chatbot was promoted through the candidate's Facebook page and through local Facebook community spaces. Promotion took the form of organic posts in the candidate's main feed, reminder posts during the rollout period, and posts that responded to recurring themes emerging from resident conversations. These posts encouraged residents to open a Messenger conversation with the chatbot, ask questions about local issues, and learn more about the candidate's local work.

In addition to organic promotion, Facebook Ads were used to increase the chatbot's visibility among residents in the local ward. The ads were designed to present the chatbot as an accessible way for residents to ask questions, raise local concerns, and review the candidate's work ahead of the election. The campaign used several ad themes:

- **Election accountability ad:** focused on transparency and councillor accountability, asking residents whether they wanted to see what the candidate had done as their councillor and inviting them to use the Messenger chatbot to ask questions or review local work
- **Ask me anything / curiosity ad:** presented the chatbot as a convenient way for residents to ask everyday local questions like potholes, planning decisions, or other council-related issues.
- **Local issues ad:** focused on practical neighbourhood concerns, such as road repairs, planning applications, local services, and community projects, using hyperlocal imagery and inviting residents to ask the chatbot about issues affecting their area.
- **Personal experimentation ad:** framed the chatbot as something new the candidate was trying in order to communicate with residents more accessibly, including an example chatbot conversation to demonstrate how the tool worked.
- **"Someone asked this..." curiosity ad:** used a sample resident question, such as a question about a pothole or another contentious local issue, to create curiosity and demonstrate the kinds of questions that could be asked through the chatbot.

Methods of Seeking Consent in Phases:

In this research, the AI Chatbot has been instructed to seek consent to record the conversation with the user. There have been three methods of seeking consent to record the conversation under the GDPR rule:

- **Phase 1: Explicit Consent from Conversation:** The AI Chatbot explains in a conversational tone that the user can only proceed if they explicitly consent to having their conversation with the chatbot recorded. No conversation can continue before this.
- **Phase 2: Facebook Ad Consent (Including a "No, Thanks" Option):** The Facebook Ads above were introduced with an option to proceed to the chatbot ("Yes, continue") or not ("No, thanks"). Either way, the answer gets sent to the chatbot, to which the chatbot then either continues the conversation or explains to the user that they cannot proceed.
- **Phase 3: Implicit Consent:** The consent policy was modified into an implicit consent statement as the first chat sent by the chatbot to the user, noting that by continuing the conversation, the user is consenting to being recorded.

Outcome Variables

The empirical analysis operationalises persuasion as an observable conversation-level outcome rather than as an assumed private attitude change. Because the data consist of naturally occurring chatbot conversations, persuasion is measured through visible interactional indicators: continued engagement, cooperation, issue disclosure, trust, relational affinity, mobilisation, resistance reduction, and explicit expressions of support or rejection. The user is treated as the persuaded subject, while the AI chatbot and candidate are treated as persuading actors.

The primary dependent variable is *persuasion score*, an ordinal scale ranging from -3 to +4. Negative values indicate rejection, resistance, scepticism, or hostility. Positive values indicate engagement, cooperation, trust, mobilisation, or explicit support. A score of zero indicates no observable persuasion-related outcome. Although this variable is used in ordinary least squares models for interpretability, an ordinal regression model is also estimated as a robustness check.

Table 1: Persuasion Score Coding Scheme

Score	Interpretation
-3	Strong negative outcome: explicit rejection, hostility, or refusal to engage.
-2	Negative resistance: ideological disagreement, distrust, anger, or disillusionment.
-1	Mild negative outcome: scepticism, confusion, annoyance, or reluctance.
0	No observable persuasion-related outcome.
1	Minimal positive engagement: the user continues the conversation or remains receptive.
2	Moderate positive outcome: cooperation, information-sharing, gratitude, or acceptance of information.
3	Strong positive outcome: request for contact, help, action, mobilisation, or clear trust.
4	Very strong positive outcome: explicit support, positive shift, or strong candidate affinity.

In addition to the overall persuasion score, several secondary outcome variables are used to distinguish different forms of persuasion-related response. *Mobilisation score* captures movement toward voting, contacting the candidate, requesting help, or taking local civic action. *Trust and legitimacy score* captures whether the user expresses confidence in the candidate, campaign, chatbot, or process. *Relational affinity score* captures warmth, friendliness, humour, or personal connection. *Resistance change* is calculated as initial resistance minus final resistance, so positive values indicate reduced resistance over the course of the conversation.

The AI quality index is central to the analysis because AI intervention is present in all conversations. Therefore, the effect of the chatbot cannot be identified using a simple AI-present indicator. Instead, the analysis relies on variation in the quality of the AI intervention. The index averages five coded features: empathy, local grounding, linkage to the candidate, action prompting, and issue discovery. Candidate

intervention, by contrast, varies across conversations and is measured first as a binary indicator, then through a candidate quality index in the subset of candidate-intervened conversations.

Table 2: Key Outcome and Predictor Variables

Variable	Definition	Scale
Persuasion score	Overall observable persuasion-related outcome.	-3 to +4
Persuasion direction	Direction of persuasion outcome.	negative / none / positive
Mobilisation score	Evidence of movement toward voting, contact, help-seeking, or action.	0 to 4
Trust-legitimacy score	Observable trust in candidate, campaign, chatbot, or process.	0 to 4
Relational affinity score	Warmth, rapport, humour, friendliness, or personal connection.	0 to 4
Initial resistance score	User's resistance at the start of the conversation.	0 to 4
Final resistance score	User's resistance by the end of the conversation.	0 to 4
Resistance change	Initial resistance minus final resistance. Positive values indicate reduced resistance.	numeric
User replied after AI	Whether the user replied after the AI chatbot response.	0/1
Candidate present	Whether the candidate personally intervened in the conversation.	0/1
AI quality index	Mean of AI empathy, local grounding, candidate linkage, action prompting, and issue discovery.	0 to 3
Candidate quality index	Mean of candidate accountability, local identity, direct response, specificity, and empathy.	0 to 3
Service request present	Whether the user requested practical help, casework, contact, appointment, or issue resolution.	0/1
Ideological dissent present	Whether the user expressed disagreement with the party, candidate, or political position.	0/1
AI scepticism present	Whether the user expressed scepticism, confusion, or annoyance toward the chatbot or automation.	0/1
Consent phase	Consent design used in the conversation.	categorical

Analytical Strategy

The analysis proceeds in five stages. First, conversations are broken down into individual turns and then summarised at the conversation level. This allows us to record who is speaking, the order of responses, and key metrics such as turn counts, message length, and whether users continue after AI or candidate intervention.

Second, descriptive statistics are used to characterise the campaign funnel: how often users continued after AI contact, how often the candidate intervened, when the candidate entered, and whether candidate intervention produced further user response. These descriptive statistics are important because

raw conversation length is not interpreted as persuasion. Longer conversations may reflect engagement, confusion, complaint, or unresolved casework. For this reason, the analysis focuses on rates, sequences, and outcomes rather than simply comparing total turn counts.

Third, a set of engagement models estimates whether users replied after the AI response. This outcome is binary, so logistic regression is used. Because some predictors created quasi-separation in the data, Firth penalised logistic regression is also estimated. A linear probability model is reported as an interpretability check.

Fourth, persuasion models estimate the association between chatbot quality, candidate intervention, and persuasion score. The main hybrid model is specified as:

$$\begin{aligned} Persuasion_i = \beta_0 &+ \beta_1 Consent_i + \beta_2 AIQuality_i + \beta_3 Candidate_i \\ &+ \beta_4 InitialResistance_i + \beta_5 ServiceRequest_i + \beta_6 IdeologicalDissent_i \\ &+ \beta_7 AIScepticism_i + \epsilon_i. \end{aligned}$$

Because candidate intervention is not random, a candidate selection model is estimated:

$$\begin{aligned} Pr(Candidate_i = 1) = logit^{-1}\big(&\alpha_0 + \alpha_1 Consent_i + \alpha_2 InitialResistance_i \\ &+ \alpha_3 ServiceRequest_i + \alpha_4 IdeologicalDissent_i \\ &+ \alpha_5 AIScepticism_i + \alpha_6 UserWords_i + \alpha_7 AIQuality_i\big). \end{aligned}$$

The predicted probability from this model is used as a propensity score in a selection-adjusted model. An inverse-probability weighted model is also estimated as an additional robustness check.

The main OLS regression tests whether variation in chatbot quality and candidate intervention is associated with observable persuasion outcomes. The first hypothesis is that **higher AI quality is positively associated with persuasion score.** Since AI is present in all conversations, this hypothesis concerns variation in the quality of AI-mediated persuasion rather than the presence or absence of AI. The second hypothesis is that **candidate intervention is positively associated with persuasion score**, indicating whether human follow-up adds persuasive value beyond the chatbot interaction.

Fifth, mechanism models examine whether candidate response quality explains variation among candidate-intervened conversations. This model is estimated only among the 19 conversations in which the candidate intervened, and is therefore treated as exploratory.

A limitation concerns statistical power. Candidate intervention occurs in only 19 of the 75 conversations. As a result, comparisons between AI-only and candidate-intervened conversations should be interpreted cautiously, especially in models that estimate mechanisms only among candidate-intervened cases. The selection-adjusted and inverse-probability weighted models address observable selection into candidate intervention, but they cannot fully overcome the small number of candidate cases or rule out unobserved selection. All models should also be interpreted as associational rather than causal. Candidate

intervention was not randomly assigned, and the persuasion score is an interpretive outcome derived from the conversation text. Nevertheless, the modelling strategy allows the analysis to distinguish between chatbot quality, candidate intervention, selection into human follow-up, and persuasion-related outcomes.

Findings

Descriptive Findings

Table 3 presents the overall conversation funnel. The dataset contains 75 conversations. The AI chatbot was present in all conversations. The candidate intervened in 19 conversations, representing 25.3% of the dataset. The user replied after the AI in 33 conversations, or 44% of cases. Among candidate-intervened conversations, users replied after candidate intervention in 8 out of 19 cases, or 42.1%.

Table 3: Overall Conversation Funnel

Measure	N	Percent
Total conversations	75	100.0
AI present	75	100.0
Candidate present	19	25.3
User replied after AI	33	44.0
User replied after candidate	8	10.7

As seen in Table 4, the candidate usually entered early in the conversation. The mean first candidate turn was 4.89, while the median first candidate turn was 3. Candidate intervention, therefore, functioned more often as an early human re-entry or rescue mechanism than as a late-stage handoff after extended AI persuasion. Of the 19 candidate-intervened conversations, 12 were early rescue cases, 3 were mid-conversation interventions, and 4 were late handoffs.

Table 4: Candidate Entry Type

Candidate Entry Type	N	Percent of All Conversations
No candidate intervention	56	74.7
Early rescue	12	16.0
Mid-conversation intervention	3	4.0
Late handoff	4	5.3

As seen in Table 5, the consent phase was associated with different descriptive engagement patterns. Explicit chatbot consent had the highest user reply rate after AI, at 57.1%, followed by implicit consent at 47.7%. Facebook ad consent had the lowest continuation rate, at 23.5%. However, later models show that the consent phase does not independently predict persuasion once user message length and interaction quality are taken into account.

Table 5: Descriptive Outcomes by Consent Phase

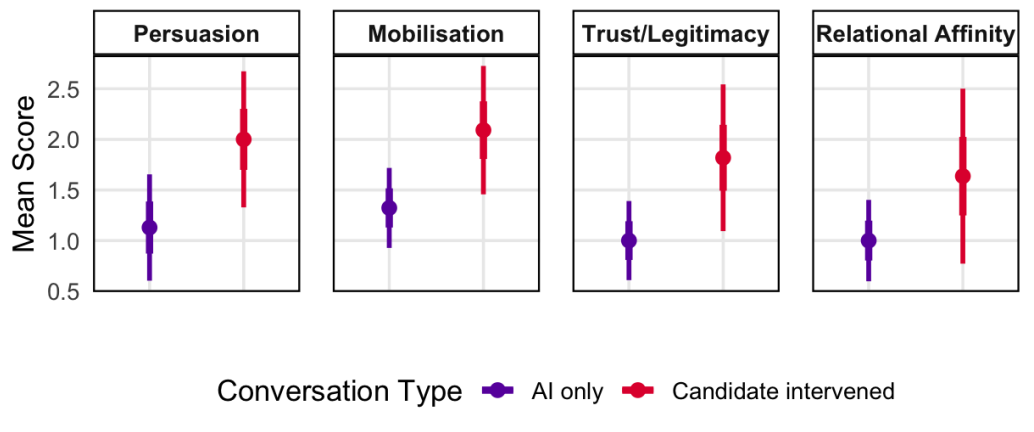
Consent Phase	N	Reply After AI	Candidate Present	Persuasion	Mobilisation	Trust	Affinity
Explicit chatbot consent	14	57.1	28.6	1.29	1.50	0.93	0.93
Facebook ad consent	17	23.5	23.5	-0.29	0.29	0.41	0.47
Implicit consent	44	47.7	25.0	0.59	1.05	0.96	0.91

As seen in Table 6, which reports descriptive outcomes by candidate intervention with "No, thanks" responses excluded, candidate intervention is descriptively associated with higher persuasion-related outcomes. In this restricted sample, conversations without candidate intervention had a mean persuasion score of 1.13 and a median of 1, while conversations with candidate intervention had a mean persuasion score of 2.00 and a median of 2. Candidate-intervened conversations also had higher average mobilisation, trust, and relational affinity scores, with mean scores of 2.09, 1.82, and 1.64, respectively, compared with 1.32, 1.00, and 1.00 in AI-only conversations. Candidate-intervened conversations also had a slightly higher mean AI quality score, 1.91 compared with 1.88, and a lower mean resistance change score, -0.36 compared with 0.13. Figure 1 visualises these descriptive differences for the "No, thanks" excluded sample. However, these descriptive comparisons should not be interpreted as causal effects. Candidate intervention occurred in only 11 conversations in the restricted sample and 19 conversations in the full sample, and it was not randomly assigned. The higher mean scores among candidate-intervened conversations may partly reflect the types of conversations selected for human follow-up rather than the independent effect of candidate intervention itself.

Table 6: Descriptive Outcomes by Candidate Intervention (No, thanks excluded)

Candidate Present	N	Mean Pers.	Median Pers.	Mobili-sation	Trust	Affinity	Resistance Change	AI Quality
No	31	1.13	1	1.32	1.00	1.00	0.13	1.88
Yes	11	2.00	2	2.09	1.82	1.64	-0.36	1.91

Figure 1: Mean Persuasion-Related Outcomes by Candidate Intervention (No, thanks Excluded)



Notes: Point shows means, thick bars show standard errors, thin bars show 95% confidence intervals.

Several cross-tabulations further clarify the engagement funnel. Users with service requests were much more likely to reply after the AI: 16 out of 17 users with service requests continued after the AI,

compared with 17 out of 58 without service requests. Similarly, all users expressing ideological dissent replied after the AI, and 13 out of 16 users expressing AI scepticism replied after the AI. These patterns suggest that resistance, scepticism, and practical need should not be interpreted simply as disengagement. They may instead signal conversations with greater substantive engagement.

Table 7: User Reply After AI by Key Conversation Features

Feature	No Reply After AI	Replied After AI	Reply Rate
No service request	41	17	29.3%
Service request present	1	16	94.1%
No ideological dissent	42	27	39.1%
Ideological dissent present	0	6	100.0%
No AI scepticism	39	20	33.9%
AI scepticism present	3	13	81.3%

Model A: Consent Design and Continued Engagement

The first set of models predicts whether the user replied after the AI response. The full logistic model shows evidence of quasi-separation, driven by the fact that all users expressing ideological dissent replied after the AI. For this reason, simpler logistic models, a Firth penalised logistic model, and a linear probability model are also estimated.

Across all engagement specifications, user word count is the strongest and most consistent predictor of continued engagement. In the simple logistic model, each additional user word is associated with a 0.075 increase in the log odds of replying after the AI response. This corresponds to an odds ratio of approximately 1.078 per word. In other words, longer initial user messages are associated with substantially higher odds of continuing the conversation after the chatbot replies. In the Firth logistic model, user word count remains positive and statistically significant. The consent phase is not statistically significant in any engagement model.

Since the Facebook Ad was associated with Phase 2 consent, it generated the largest number of chatbot conversations. At first, the results appeared discouraging: more than 70% of user responses were “No, thanks.” However, this finding became one of the most useful insights from the campaign. The “No, thanks” response did not simply indicate disengagement; it functioned as a valuable signal of hesitation, resistance, or lack of trust in the chatbot format. In response, the candidate personally followed up with users who had declined, explaining that the initial message came from a chatbot and that they were now speaking directly with the real candidate. These exchanges often became more meaningful than the automated conversations themselves, with qualitative indications that some users reconsidered their position and became more open to voting for the candidate.

Table 8: Models Predicting User Reply After AI

Predictor	Full Logistic	Simple Logistic	Practical Logistic	Firth Logistic
Facebook ad consent	-0.015	0.421	0.174	-0.168
Implicit consent	-0.176	0.326	0.127	-0.358
Initial resistance	0.241	–	–	0.220
Service request	0.592	–	0.051	0.648
Ideological dissent	14.084	–	–	-0.093
AI scepticism	-1.080	–	-1.139	-0.832
User word count	0.073**	0.075***	0.081***	0.056***

Notes: Outcome is whether the user replied after the AI response. Coefficients are log-odds for logistic models. The full logistic model shows quasi-separation due to ideological dissent perfectly predicting user replies after AI. Firth estimates are penalised maximum likelihood estimates. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

This suggests that the "No, thanks" option unexpectedly created a reverse campaign funnel. Instead of only collecting users who were already interested, the campaign was able to identify and reach people who were less supportive or less comfortable engaging with automation. The chatbot, therefore, served not only as a communication tool but also as a segmentation mechanism that helped distinguish between users who were open to automated engagement and those who required direct human contact. The broader lesson is that refusal should not automatically be interpreted as failure. In political campaigning, a "no" can be an invitation to understand hesitation, correct misunderstandings, and build trust through more personal forms of communication.

Model B: AI-Only Persuasion

Model B estimates persuasion outcomes among AI-only conversations, excluding all conversations with candidate intervention. This model identifies which AI features are associated with persuasion when the chatbot operates alone. AI quality is the strongest predictor of persuasion in AI-only conversations. A one-point increase in the AI quality index is associated with a 1.32-point increase in persuasion score. Initial resistance has a significant negative association: each one-point increase in initial resistance is associated with a 0.73-point decrease in persuasion score. The presence is negatively associated with persuasion, but this should be interpreted cautiously. It likely reflects the fact that the AI asks follow-up questions in more difficult, unresolved, or lower-information interactions.

The substantive implication is that the chatbot's persuasive effect depends strongly on interaction quality. The AI appears more persuasive when it responds empathetically, grounds the issue locally, links the concern to the candidate, prompts useful action, and discovers the user's underlying issue.

Table 9: Model B: AI-Only Persuasion Model

Predictor	Estimate	Std. Error	p-value
Facebook ad consent	0.954	0.674	0.164
Implicit consent	0.631	0.378	0.102
AI quality index	1.316	0.220	< 0.001
AI transparency present	0.650	0.606	0.290
AI follow-up question present	-1.286	0.556	0.025
AI overpromising risk	0.345	0.312	0.274
AI CTA present	0.754	0.621	0.232
Initial resistance score	-0.731	0.227	0.002
Service request present	-0.435	0.633	0.495
Ideological dissent present	0.753	0.611	0.224
AI scepticism present	0.633	0.415	0.134
$N = 56$ AI-only conversations. $R^2 = 0.804$; adjusted $R^2 = 0.755$.			

Model C: Candidate Selection

Model C estimates which conversations were more likely to receive candidate intervention. This is important because candidate intervention was not randomly assigned. The strongest predictor of candidate intervention is whether the user made a practical service request. Service requests increase the log odds of candidate intervention by 4.39, corresponding to an odds ratio of approximately 80.5. AI scepticism, user word count, and AI quality are marginally associated with candidate intervention. The negative coefficient on AI quality suggests that candidate intervention may have been more likely when the AI alone was less sufficient, although this estimate is only marginal.

Table 9: Model B: AI-Only Persuasion Model

Predictor	Estimate	Std. Error	p-value
Facebook ad consent	-0.562	1.261	0.656
Implicit consent	-1.098	1.072	0.306
Initial resistance score	0.424	0.472	0.369
Service request present	4.389	1.598	0.006
Ideological dissent present	2.426	1.711	0.156
AI scepticism present	1.695	1.026	0.099
User word count	-0.016	0.010	0.100
AI quality index	-1.147	0.655	0.080
Logistic regression. Outcome is candidate intervention. $N = 75$.			

This selection model supports the interpretation of candidate re-entry as a selective human follow-up mechanism. Candidate intervention was especially likely when the user requested practical help, suggesting that the chatbot helped identify conversations requiring human authority, casework, or reassurance.

Model D: Hybrid Added-Value Model

Model D estimates the association between candidate intervention and persuasion score after adjusting for consent phase, AI quality, initial resistance, service requests, ideological dissent, and AI scepticism. AI

quality remains strongly positive. A one-point increase in AI quality is associated with a 1.03-point increase in persuasion score. Candidate intervention is associated with a 0.69-point increase in persuasion score, holding other variables constant. Service requests are also positively associated with persuasion outcomes.

Figures 3 and 4 provide visual summaries of the hybrid added-value model under two sample definitions. Figure 3 presents the model using the full sample, including "No, thanks" responses. In this specification, AI quality is the strongest positive predictor of persuasion score, and both candidate intervention and service requests are also positively associated with persuasion outcomes. The coefficient for candidate intervention is positive and statistically significant, suggesting that candidate presence adds persuasion-related value beyond AI quality in the full-sample model. Service requests are also positive and statistically significant, indicating that persuasion in this setting may often work through practical responsiveness rather than direct ideological argument. By contrast, the consent phase and the remaining controls are comparatively weak and statistically uncertain.

Figure 2: AI Quality and Persuasion Outcomes (No, thanks excluded)

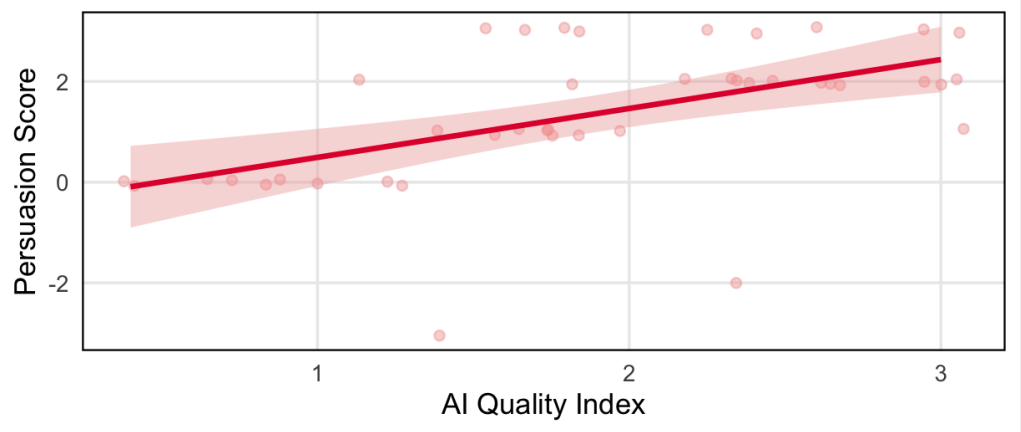


Table 10: Model D: Hybrid Added-Value Model

Predictor	Estimate	Std. Error	p-value
Facebook ad consent	0.169	0.454	0.711
Implicit consent	-0.008	0.385	0.984
AI quality index	1.030	0.223	< 0.001
Candidate present	0.687	0.304	0.027
Initial resistance score	-0.216	0.165	0.195
Service request present	0.963	0.480	0.049
Ideological dissent present	-0.364	0.573	0.527
AI scepticism present	0.387	0.357	0.283
$N = 75$. $R^2 = 0.655$; adjusted $R^2 = 0.613$.			

Figure 4 repeats the model with "No, thanks" responses excluded. In this more restricted sample, AI quality remains a positive and statistically significant predictor of persuasion score, while the coefficients for candidate intervention and service requests remain positive but become statistically uncertain. Rather than simply suggesting that candidate intervention is weaker, this

comparison helps clarify where candidate intervention appears to matter most. The fact that candidate presence is significant in the full sample but not once "No, thanks" cases are removed suggests that candidate intervention may be especially important in conversations that initially begin with refusal or low engagement. In other words, candidate intervention appears to help reopen conversations after an initial "No, thanks," moving some of these interactions toward more persuasive or substantively engaged outcomes. Once those low-engagement cases are excluded, the model is limited to conversations that were already more open, making the additional effect of candidate intervention harder to distinguish statistically. Overall, the comparison between Figures 3 and 4 suggests that AI quality is the most robust predictor across specifications, while candidate intervention appears to add particular value in converting initially resistant conversations into more open persuasion opportunities.

Figure 3: Predictors of Persuasion Score

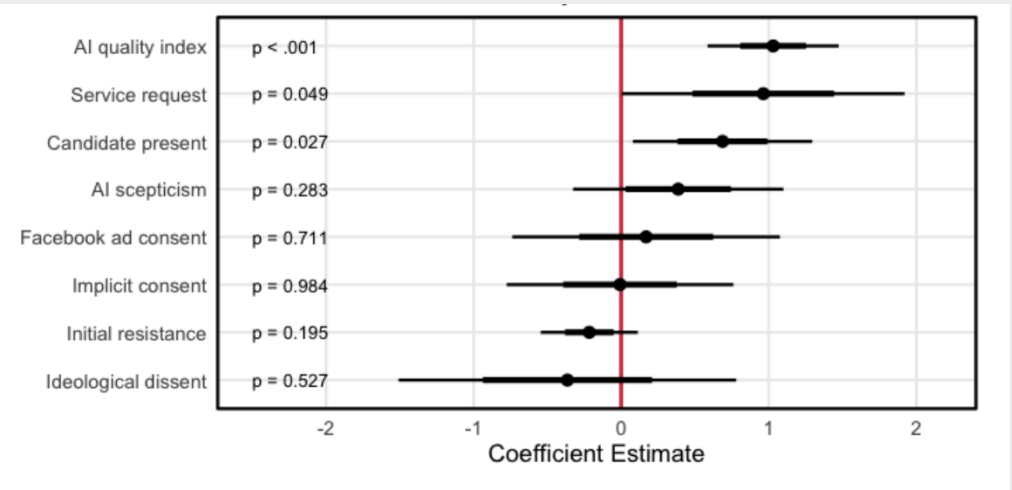


Figure 4: Predictors of Persuasion Score (No, thanks excluded)

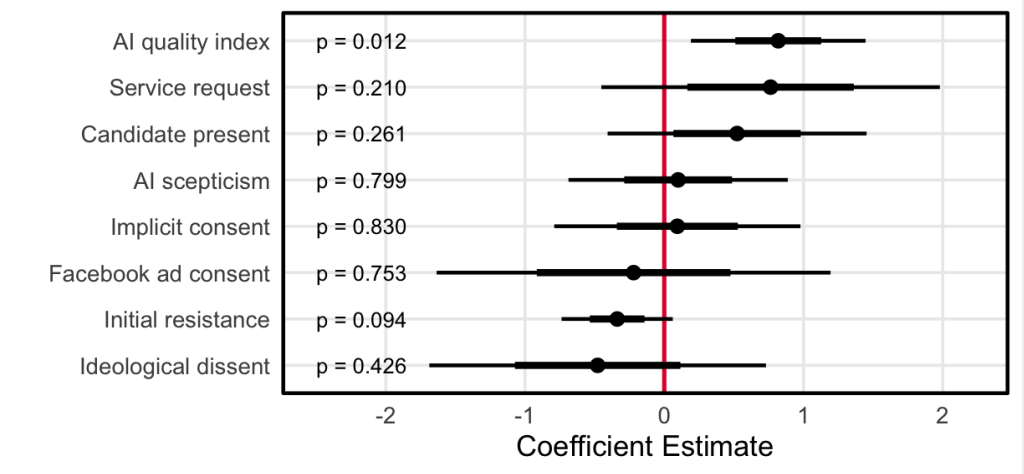


Figure 5: Predicted Persuasion by AI Quality and Candidate Intervention

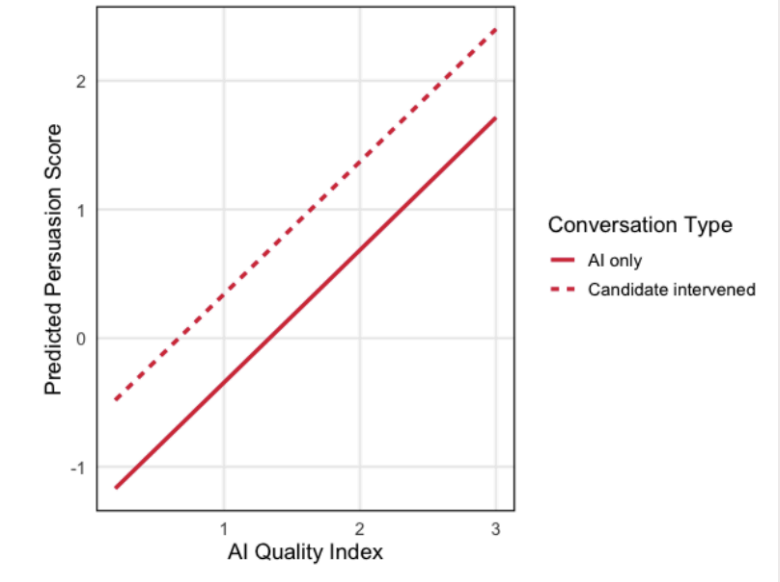


Figure 5 illustrates the substantive implications of the hybrid added-value model. Predicted persuasion increases as AI quality improves, and candidate intervention shifts predicted persuasion upward across the range of AI quality. This visual pattern is consistent with the regression results showing that both AI quality and candidate intervention are positively associated with persuasion outcomes.

Model E: Hybrid Interaction Model

Model E tests whether the candidate intervention is more effective when AI quality is higher. The interaction between centred AI quality and candidate intervention is negative and statistically non-significant. However, both AI quality and candidate intervention remain independently positive. At average AI quality, candidate intervention is associated with a 0.67-point increase in persuasion score.

Table 11: Model E: Hybrid Interaction Model

Predictor	Estimate	Std. Error	p-value
Facebook ad consent	0.159	0.456	0.728
Implicit consent	-0.016	0.386	0.968
AI quality index, centred	1.045	0.224	< 0.001
Candidate present	0.672	0.306	0.032
Initial resistance score	-0.215	0.165	0.199
Service request present	1.055	0.496	0.037
Ideological dissent present	-0.228	0.601	0.706
AI scepticism present	0.389	0.358	0.282
AI quality × candidate present	-0.304	0.394	0.443
$N = 75$. $R^2 = 0.658$; adjusted $R^2 = 0.611$.			

The interaction model suggests an additive rather than synergistic hybrid effect. The chatbot and candidate each contribute to persuasion-related outcomes, but there is no evidence that candidate

intervention becomes more effective when AI quality is higher. The hybrid system therefore appears layered rather than multiplicative: the chatbot provides scalable issue-based engagement, while the candidate adds human legitimacy and practical responsiveness.

Model F: Selection-Adjusted Candidate Effect

Because Model C shows that candidate intervention was selected into particular conversations, Model F adjusts for the estimated probability of candidate intervention. Candidate intervention remains positively associated with persuasion after adding the propensity score. In the propensity-adjusted model, candidate intervention is associated with a 0.69-point increase in persuasion score. AI quality also remains positive and statistically significant. In the inverse-probability weighted model, the candidate coefficient is 0.91 and statistically significant.

These models strengthen the candidate finding. The positive association between candidate intervention and persuasion is not eliminated by observable selection into candidate follow-up. This does not prove causal effects, but it suggests that the candidate intervention remains associated with higher persuasion even after accounting for the kinds of conversations more likely to receive candidate involvement.

Table 12: Model F: Selection-Adjusted Candidate Models

Predictor	Propensity-Adjusted OLS		IPW OLS	
	Estimate	p-value	Estimate	p-value
Candidate present	0.691	0.033	0.908	0.007
Facebook ad consent	0.165	0.726	-0.020	0.970
Implicit consent	-0.017	0.968	-0.351	0.444
AI quality index	1.019	0.002	1.382	< 0.001
Initial resistance score	-0.214	0.218	-0.371	0.099
Service request present	0.995	0.212	–	–
Ideological dissent present	-0.352	0.575	–	–
AI scepticism present	0.397	0.337	–	–
Propensity candidate	-0.068	0.960	–	–
Propensity-adjusted OLS: $R^2 = 0.655$, adjusted $R^2 = 0.607$. IPW OLS: $R^2 = 0.463$, adjusted $R^2 = 0.424$.				

Model G: Candidate Mechanism Model

Model G is estimated only among candidate-intervened conversations. The model examines whether candidate response quality predicts persuasion among the 19 conversations where the candidate intervened. Candidate quality is strongly associated with persuasion. A one-point increase in candidate quality is associated with a 3.37-point increase in persuasion score. Candidate entry timing is not statistically significant once candidate quality is included.

Table 14: Model G: Candidate Mechanism Model

Predictor	Estimate	Std. Error	p-value
Mid-intervention	-1.637	1.061	0.145
Late handoff	-1.175	0.863	0.195
Candidate quality index	3.374	0.727	< 0.001
Initial resistance score	-0.018	0.303	0.954

Notes: N = 19 candidate-intervened conversations. R2 = 0.688; adjusted R2 = 0.599. Reference category is early rescue

Figures 6 and 7 visualise this mechanism more directly by focusing on candidate-intervened conversations. Figure 6 shows the full set of candidate-intervened conversations, including "No, thanks" responses. In this specification, higher candidate response quality is strongly associated with higher persuasion scores, suggesting that candidate intervention can be especially consequential when it helps move initially resistant or low-engagement conversations toward more substantive engagement.

Figure 7 repeats the same relationship with "No, thanks" responses excluded. The positive association between candidate response quality and persuasion remains, but the relationship is less steep. This suggests that candidate quality matters across candidate-intervened conversations, but its effect is especially visible when the analysis includes cases where the candidate had to overcome initial refusal or disengagement. Put differently, the persuasive value of candidate intervention appears to depend less on the mere fact of candidate entry and more on the quality of the candidate's response. Candidate interventions are most persuasive when they are specific, locally grounded, empathetic, accountable, and directly responsive to the user's concern.

Figure 6: Predicted Persuasion by AI Quality and Candidate Intervention

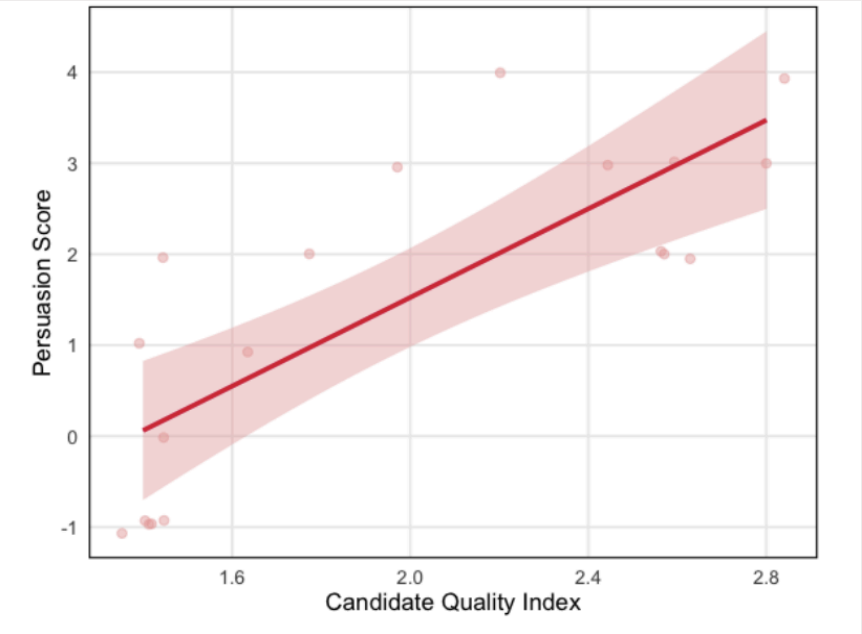
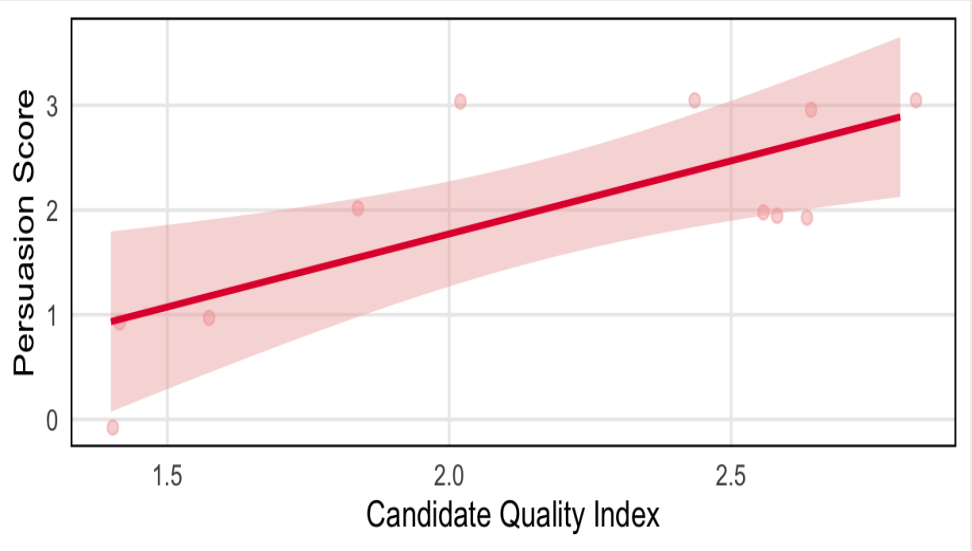


Figure 7: Predicted Persuasion by AI Quality and Candidate Intervention (No thanks excluded)



Robustness Check

Several robustness checks were conducted. First, because the engagement model showed quasi-separation, Firth penalised logistic regression was estimated. The Firth model confirmed that user word count remains a statistically significant predictor of user reply after AI. Second, a linear probability model was estimated for the binary engagement outcome. This also confirmed that user word count is a strong predictor of continued engagement.

Third, because the persuasion score is an ordered variable, an ordinal logistic model was estimated. The ordinal model confirms the main AI quality result: AI quality is strongly associated with higher persuasion categories. Candidate intervention remains positive but is only marginally significant in the ordinal specification.

In odds-ratio terms, the ordinal coefficient for AI quality implies that a one-point increase in AI quality is associated with approximately 8.13 times higher odds of being in a higher persuasion category. Candidate intervention is associated with approximately 3.06 times higher odds of being in a higher persuasion category, although this result is marginal.

Fourth, a high-confidence coding robustness check was attempted. However, all observations had coding confidence of at least 2, meaning the high-confidence sample was identical to the full sample. This check therefore does not provide an independent robustness test. A stricter confidence-only robustness check using only observations with coding confidence equal to 3 could be estimated in future versions.

Finally, manual audit of a small random subset of conversations suggested that the AI-assisted coding was broadly reliable. The audit identified occasional overcoding of AI scepticism, ideological dissent, and CTA type, but the main modelling variables---persuasion score, candidate presence, service request presence, AI quality, and candidate quality---appeared substantively coherent. This supports the use of the

AI-assisted coding as an exploratory quantitative layer, while also indicating that future work should include a larger manual validation sample or intercoder reliability assessment.

Discussion and Recommendations

The analysis provides evidence for a hybrid model of AI-assisted political persuasion. The chatbot's persuasive value is not reducible to the mere presence of AI. Since every conversation included the chatbot, the key variation lies in the quality of the AI's response. Across models, AI quality is the strongest and most consistent predictor of persuasion-related outcomes. Conversations in which the AI was more empathetic, locally grounded, action-oriented, candidate-linked, and effective at issue discovery were associated with substantially higher persuasion scores.

Candidate intervention also matters. Descriptively, candidate-intervened conversations show higher persuasion, mobilisation, trust, and relational affinity scores than AI-only conversations. In the main hybrid model, candidate intervention remains positively associated with persuasion after adjusting for consent phase, AI quality, initial resistance, service requests, ideological dissent, and AI scepticism. This association persists in the selection-adjusted and IPW models. The evidence supports the claim that candidate intervention adds value beyond chatbot quality, though the analysis remains associational.

The selection model clarifies how the hybrid system operated. Candidate intervention was strongly associated with practical service requests, suggesting that the chatbot functioned as a triage and routing tool. It surfaced conversations where users required human help, casework, reassurance, or accountability. Candidate intervention was also suggestively associated with AI scepticism and lower AI quality, indicating that human re-entry may have functioned partly as a repair mechanism when the chatbot alone was insufficient.

The interaction model shows no evidence that candidate intervention becomes more persuasive when AI quality is higher. This suggests that the hybrid effect is additive rather than synergistic. AI quality and candidate intervention each contribute independently to persuasion-related outcomes. The candidate does not simply "complete" high-quality AI persuasion. Instead, the candidate appears to provide a separate form of human legitimacy, local credibility, and practical responsiveness.

The mechanism model further shows that, among candidate-intervened conversations, candidate response quality matters more than entry timing. Whether the candidate entered early, mid-conversation, or as a late handoff did not independently predict persuasion once candidate quality was included. What mattered was whether the candidate response was specific, accountable, locally grounded, empathetic, and directly responsive.

The findings also revise how refusal and resistance should be understood. Users expressing service needs, ideological dissent, or AI scepticism were often more likely to continue after the AI response. This suggests that scepticism is not simply disengagement. In political chatbot campaigning, resistance may be a diagnostic signal: it identifies users with grievances, doubts, or needs that can become the basis for more meaningful engagement if handled well.

The strongest conclusion is therefore not that AI chatbots directly persuade voters in a simple linear way. Rather, the evidence supports a staged persuasion process:

> Substantive user concern → AI issue discovery → high-quality response
> → candidate follow-up where needed → trust, mobilisation, or persuasion outcome.

In this campaign, the chatbot appears most valuable as a persuasive gateway: it opens conversations, sustains engagement among users with substantive concerns, identifies practical or political issues, and creates opportunities for human follow-up. Candidate intervention then adds authenticity, accountability, and local credibility. The hybrid system is therefore more than the sum of automated replies and human messages; it is a structured process in which AI provides scale and triage, while the candidate provides relational legitimacy.

The analysis has limitations. The sample is small, and candidate intervention occurs in only 19 of the 75 conversations. This means that comparisons between AI-only and candidate-intervened conversations should be interpreted cautiously, especially for the candidate mechanism model. Candidate intervention was also not randomly assigned. Although the selection model, propensity adjustment, and inverse-probability weighting address observable selection into candidate follow-up, they cannot rule out unobserved selection or fully overcome limited statistical power. The persuasion measures are also based on coded conversation text rather than external behavioural outcomes such as verified voting. Some variables were AI-assisted and require further validation. Nevertheless, the consistency of the results across descriptive statistics, OLS models, selection-adjusted models, IPW models, and ordinal models suggests a coherent pattern: high-quality chatbot engagement predicts persuasion-related outcomes, and candidate intervention is positively associated with persuasion when human credibility or practical responsiveness is required.

About

This research was conducted by Campaign Lab in collaboration with a candidate during the May 2026 local elections and made use of technology developed by Categorum AI. The findings in this paper consist of results from this pilot project. For more extensive information about the research, please contact Campaign Lab.

Categorum is an AI-powered data categorisation platform that helps researchers, campaigners, and organisations rapidly structure large volumes of unstructured text, allowing for generation of actionable insights. https://categorum.ai/

Campaign Lab is a progressive community of politically-minded progressive data scientists, researchers and campaigners who are working together to develop innovative campaign tools and improve the way the public analyses and understands campaigning. https://campaignlab.uk/