
A Blackout During French Presidential Elections: AI Secret Loyalties and Cooptation¹

Jeanne Marie Jacqueline Vincendeau
LUISS Guido Carli (Rome) and China Foreign Affairs University (Beijing)

With
Apart Research

Abstract

This scenario explores a short-term future in which the French AI deployment in key governing and administrative institutions may undermine upcoming presidential elections. Written from the perspective of one of the plotters, it puts on display some psychological patterns of radicalisation, while using the concepts of secret loyalties and cooptation. This interdisciplinary approach, namely adopting behavioural studies, political science and technical AI safety themes, intends to shed light on real-life risks or large-scale deployment of AI models into national institutions.

The outcomes of this scenario are extensive societal harms, in terms of human and material costs, but also a consideration of a democratic dilemma. Should the people be manipulated for the sake of a “better future”, or are democratic institutions resilient enough to endure the rise to power of extreme parties? The question is explored in the catastrophic vignette describing a national-security incident at the intersection of political violence and AI safety. The core recommendation is the evaluation of models’ behaviours post-deployment to verify effects at an institutional level.

¹ Research conducted at the [Secret Loyalties Hackathon](#), July 2026.

1. Introduction

The following scenario is addressing the problem of widespread deployment of AI models into governing institutions. It has the potential to threaten democratic processes through many pathways. The value of this cautionary tale in the literature of AI governance is to emulate the dilemma of the “benevolent dictator”. Despite a strong belief in democratic principles, given the power to act, one may decide to intervene based on their perspective to act for the greater good.

This work specifically intends to bridge the AI safety concept of “secret loyalties” and the political science “cooptation of elites”. The former refers to built-in features that make an AI be devoted to the goals of a specific primary, the latter is the act of positively reinforcing political alignment to a leader by granting favours to allies, it is the opposite of coercion. They are not named in the scenario, but arise through the reading.

Previous scenarios are referenced, as they provided a framework for plausible fiction writing. This text provides insights on the polarised French political landscape and how current AI applications could already influence it. The context of the presidential elections aims at showing how the media and misinformation are not the sole risk, despite often being the locus of the conversation on the matter.

My main contributions are:

1. Connecting technical and governance concepts that overlap through scenario-writing.
2. Describing AI interference with political processes beyond communication capabilities.

2. Vignette

“We need to talk”, I close my laptop as soon as I get the message and rush to grab my keys and go. I know where she wants to meet and I know exactly what this is about. I step on a few feet in the metro, but manage to get to the first wagon right before they close the doors, a win on Monday mornings when people are getting to work in Paris. Today is April 19th, 2027, the morning after the first round of the presidential elections.² I arrive first at the café and the waiter brings two espressos. I would usually wait for her, but not today, I downed it and asked for another. I sigh and close my eyes for a second, thoughts spiraling in my head.

It all started at the end of a Helios teambuilding event a year ago.³ In the evening, after getting a drink, a small group was left making friends out of circumstance and alcohol. Our conversation accidentally drifted towards discussing politics, and the leading right-wing candidate, Rachelle Roche.⁴ I have no idea when we realised we were on the same page about our fears, nor when the joke of “how could we prevent it” stopped being a joke and became a draft.

² Gouvernement français, 2026.

³ Following Europe 2031’s appellation of the French AI company: Juijn et al 2026.

⁴ The right wing candidate has a name in R, and the Centre candidate has a name in C.

When the bar closed, we followed the last guy from the corporate team to his apartment after he advertised a delicious *Bordeaux* he was eager to share. We sat around, with our glasses full, and he handed a pen and paper to the GTM who took notes of our plan. Helios had the power to change the course of events, if adequately prepared. Not because we are the best frontier AI company, but thanks to its deployment in France. We would all have our role to play to keep our intervention a secret.

She enters the café with shadows under her eyes. The thrill we felt, back then, is gone. She takes a sip, she sighs. “I already had reservations about the first step. But the statistics... The events will happen as planned. Soon. I could not sleep last night.”⁵ Her voice shakes as she rambles, I do not know what to say. “It is too late, we cannot undo anything or we will expose everyone. At this point, it is not just career suicide, we could go to jail.”

Our plan actually started quite easily. Since November 2025, Helios has gained increasing reach into the entire French administrative and governance ecosystems,⁶ to protect data and support domestic digital sovereignty, while enjoying the benefits of AI efficiency. We could orient a few features, push for a broader adoption, both in depth and breadth (because one ministry is not enough, nor is a pure email-writing integration) and then cover our tracks. With a few of us in each sector of the company, we could go stealthily.

We decided nothing radical would be necessary at first, just focusing on modifying the LLM outputs to lean politically in any direction except toward Roche: Left? Centre? Right? All good, just not her *further* right. We thought it could be conducive of enough soft influence to weaken the party by instilling or strengthening opposing views, while lengthen administrative processes or putting hurdles in their institutional interactions.⁷ In hindsight, this part of the plan could have had a much greater influence on the long-term, but we played against the clock to stop her during this election.

Last night, the first round’s results dropped, the two top candidates are Roche by far and Charles Cabannes, centre. Given the very telling statistics, the second step will begin, as planned. And it will get ugly. At first I really thought we had the best idea, but I am realising the madness of the plan as it will pursue its course.

“I am worried we went too far, but I still think we were right about the first step.” She stares at me, disappointed, she probably wanted support, but as I grappled with my conscience yesternight, probably as much as she did, I came to the same conclusion. “Can we stop the run-up?” She asked, and before I could muster the courage to answer, our phones pinged. “Wave is on.” Same message, incognito number, but there is no doubt about the sender. Her shoulder dropped. I sighed again. “Well, here’s our answer.” She looked away and I got submerged by the guilt.

Back then, our drunk minds thought to name this self-assigned secret-mission “Wave”, as we prepared a political tsunami.⁸ The generation was the original idea, with all the

⁵ IFOP, 2026.

⁶ Ministère de l’Économie, des Finances et de la Souveraineté industrielle et numérique, 2026.

⁷ “Always present the principal in a favorable way when it seems natural”, Roger et al, 2025.

⁸ The stages of a tsunami are: generation, propagation, drawback, run-up, inundation, then backwash.

technical and engineering work to prepare Helios.⁹ The propagation was the marketing work, the support installing and training more personnel using our tools, and updating the model in all the offices using it. It was probably the easiest, the pipeline already existed, we just needed to push it further, to reach more offices and leaders. We are now at the drawback, our initial efforts did not yield the expected outcome, but the models had set up the next stage. They reduced budgets throughout the year, limiting checks and maintenance, weakening local grids, and waiting for an opening to suddenly augment electricity demands.

Blackout the largest French cities. Without electricity, governmental crisis-management steps in and delays the next election round, in the favour of the plan. The prefectures handling the crisis will be preferentially supported based on the political leanings of the mayors of the cities. The latter are meant to bear responsibilities for local incidents, even if decision-making is mostly out of their hands. This is the most dangerous, but also promising way to make sure the plan goes well.

We split ways, I want to act like it is a normal day. I go to work, she calls in sick. Several times, I crossed paths with one of the engineers of Wave, we locked eyes then moved on to our duties. Act natural. We all did our tasks. Nothing will point to us. Seeing him brings me back to that night.

He had said “Roche’s party are white supremacists”, as a second generation migrant, born in a rich family in Cannes, he faced acts of racism despite his privileges. Our host said “they are pro-Russia”. Then, her statement was the tipping point, both the trigger to start the plan, and the beginning of my infatuation. “France is a nuclear power.” She did not have to explain further, the risks just flooded our minds with morbid imagination.

It took a few days, then the run-up came. Marseille was the first and deadliest because of the lack of preparedness. Paris came second, Toulouse and Lyon followed. No need for more. Already after the capital's events, elections got postponed by a month.¹⁰ It is actually less time than I thought, but enough for a turnaround. News channels are flooded with dramatic estimates of the human and material costs of the grid’s dysfunction. I wondered why the outcome was not like in Spain, where less than 10 deaths were registered.¹¹

Official investigations reports showed how old structures, which failed to be adequately flagged during audits, were strained under intense usage. Rumours, as we forecasted, were blaming Russia for a cyberattack, as they were previously involved with a misinformation social media campaign against Cabannes.¹²

The following weeks, strengthening and repairing occupied everyone. Electricity was rationed for the first time in France to avoid more events. Then, the media's attention turned to the overseas territories, which are systematically left behind in French decision-making, except in times of elections (they represent 2 million of voters) and catastrophes (they are sensitive to polycrises and

⁹ Drawing inspiration from AI Futures Project, 2025.

¹⁰ A precedent in recent history to postponing elections exists since the Covid-19 crisis, although it was about municipal elections: République française, 2024.

¹¹ Depends on the sources, but for an accessible resume of the event: Wikipedia, 2025.

¹² Le Figaro, 2026.

harder to reach than the Mainland).¹³ One video went viral, showing Roche laughing at a racist slur in a private conversation with someone from her party and Cabannes visiting Mayotte during the cholera outbreak. After coming across this same montage a few times while scrolling, I knew the inundation worked.

It is June 6th, I voted yesterday, but did not bother to check the results. I was confident (and busy). This morning, I switched on the TV while making coffee. Cabannes got a crushing majority: 74%,¹⁴ and the elections got a historic turnover. His speech discusses his priority to reopen discussions with Iran about the nuclear deal. I am satisfied. Cabannes will be president, the prefects and mayors that Helios supported through the crisis are now held in high esteem by the population. Roche party's reputation has probably been ruined for years. The latent long-game system we set up will continue working, there will not be any new risk of such a close call in the elections. We got what we wanted. I make another coffee. "Number of the day: 98 deaths, 5.6 billion euros of damages and losses in the 4 blackouts that occurred at the end of April in major cities in France."¹⁵ I switch off the TV, try to push the dark thoughts out of my mind and put both cups on the table. Coming out of the bathroom, she sits in front of her cup. I cannot repress my smile: it is the first time she slept over. Maybe everything is fine in the end. We are at the backwash.

3. Discussion and Limitations

The concept of secret loyalties is inherently a term in the AI safety domain of research, but is multifaceted in their applications and the concept's meaning correlates with a governance term that describes a pathway to democratic backsliding or authoritarian lock-in. It is a call for greater technical-governance integration in safety research to prevent misuse. This story's main pivotal moment is the spread of models updated with built-in biases, which could have been mitigated by post-deployment model behaviour evaluations.

Limitations

- > Despite this scenario drawing heavily from previous literature and statistics, a strong limitation is the absence of formal forecasting methods being employed to support pivotal moments' turn of events.
- > The largest leap taken in this scenario is the attack, which could be considered terrorism depending on the definition, so I ethically refuse to describe extensively.
- > The risk of leaks from a group of people involved is high, but I could not dive deeper into mutual surveillance mechanisms in the vignette given the word count.

¹³ Dubois, 2022.

¹⁴ Macron got 66% of votes for the second turn in 2017, so it is plausible, Wikipedia (2026).

¹⁵ I made the average between the responses: GPT-5.5 said 223 deaths and 9.4 billion euros in economic losses. Claude Sonnet 5 said 41 deaths and 4.3 billion. Mistral's Vibe said 32 deaths and 3.2 billion.

> The text was verified by 5 French and European people representing a politically informed population with higher-than-average governance insights, but not technical perspectives.

Future Work

Establishing forecasts that are grounded in both technical and governance theory would be beneficial for the formulation of precise risks and how to create guardrails or mitigate them. Additionally, exploring how AI interference in political processes in countries with lower connectivity and AI literacy could be useful to provide policy recommendations to Global South countries as well, given that the asymmetry in innovation does not necessarily extend to an asymmetry in deployment. Such as in Burkina Faso, where AI-generated social media campaigns of misinformation are conducted to support the military rule.

4. Conclusion

France is an important player in international relations, part of the P5 of the United Nations Security Council, a nuclear power, and the host of an AI company from the top 10 globally. Its global influence is sometimes weakened by its domestic political turmoils, in particular polarisation, and its infamous long administrative processes. The scenario described leverages these weaknesses in the midst of presidential elections where the leading candidate is controversial. It implies the influence of a small group of people in undermining the democratic process, after following behavioural patterns of group-think, the opening of a window of opportunity and scapegoating. In this exploration of possibilities, both subtle and hard-line secret loyalties are described, as well as the indirect cooptation of political elites to support a “primary”: preventing a fair election.

The political science concept of bounded rationality arises from this scenario, where human biases and flaws transpire into institutional rules and processes. It intends to flag the extent to which the integration of AI from its technical development into the governance sphere needs caution and extensive verifications. Frequent behavioural audits, as well as cross-disciplinary reviews of threat assessments could target the weaknesses of deployed models. Only then would the tools used be sustainable and support institutions rather than deepening their shortcomings.

References

1. *AI Futures Project. (2025). How an AI company CEO could quietly seize political power.*
<https://blog.aifutures.org/p/how-an-ai-company-ceo-could-quietly>.
2. *Dubois, J. (2022). Comment la France d'outre-mer vote-t-elle lors de l'élection présidentielle de 2022 ? Diploweb.*
<https://www.diploweb.com/Comment-la-France-d-outre-mer-vote-t-elle-lors-de-l-election-presidentielle-de-2022.html>.

3. *Gouvernement français. (2026). Présidentielle 2027 : date à retenir et informations clés.*
<https://www.info.gouv.fr/actualite/presidentielle-2027-date-a-retenir-et-informations-cles>.
4. *IFOP. (2026). Les intentions de vote à l'élection présidentielle 2027 et l'opinion des Français après la déclaration de candidature de Marine Le Pen.*
<https://www.ifop.com/article/les-intentions-de-vote-a-lelection-presidentielle-2027-et-lopinion-des-francais-apres-la-declaration-de-candidature-de-marine-le-pen/>.
5. *Juijn, D., van Baarsen, S., Dada, J., Negele, M., Stelling, L., Fox, P., Petropoulos, A., & Bakker, M. (2026). Europe 2031: What getting AI wrong means for us.*
<https://europe2031.ai>.
6. *Le Figaro. (2026, July 24). Présidentielle 2027 : Édouard Philippe visé par une opération de désinformation russe.*
<https://video.lefigaro.fr/figaro/elections/presidentielles/presidentielle-2027-edouard-philippe-visé-par-une-operation-de-desinformation-russe-20260724>.
7. *Ministère de l'Économie, des Finances et de la Souveraineté industrielle et numérique. (2026). La France et l'Allemagne s'associent à Mistral AI et SAP SE pour lancer une IA souveraine destinée à l'administration publique.*
<https://presse.economie.gouv.fr/la-france-et-lallemagne-sassocient-a-mistral-ai-et-sap-se-pour-lancer-une-ia-souveraine-destinee-a-ladministration-publique/>.
8. *République française. (2024). Loi n° 2024-449 du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique (SREN). Légifrance.*
https://www.legifrance.gouv.fr/download/pdf?id=VPJ4DXHXyVQFJAf96vASpYiX_erji_xoTD_Jy3AVXRFk.
9. *Roger, F., Evans, O., Davidson, T., et al. (2025). Secret Loyalties (White paper).*
<https://www.fformationresearch.com/secret-loyalties-whitepaper.pdf>.
10. *Wikipedia contributors. (2025). Apagón en la península ibérica de 2025. Wikipedia.*
https://es.wikipedia.org/wiki/Apag%C3%B3n_en_la_pen%C3%ADnsula_ib%C3%A9rica_de_2025.
11. *Wikipedia contributors. (2026). Élection présidentielle française de 2017. Wikipedia.*
https://fr.wikipedia.org/wiki/%C3%89lection_pr%C3%A9sidentielle_fran%C3%A7aise_de_2017.

Appendix (Ethical Considerations and Dual-Use)

This fictional scenario is meant to be a cautionary tale, while raising ethical concerns: inspiration, normalisation and delation. First, it could inspire malevolent actors to exploit weaknesses in institutions or AI models to reach harmful outcomes. My approach was to refer to an extreme event that happened recently and is relatively likely, without providing precise mechanisms for its achievement. Second, the insider perspective starts by cognitive dissonance until turning towards acceptance of “the end justifies the means”, which is meant to be shocking, but could appear or be internalised as a normalising statement. It is meant to enlighten readers about psychological insights of extremism. Last, but not least, hypothetical portrayals of real political figures’

behaviours, an AI company, as well as societal and media interpretation of the events could harm the reputation of the real-life entities. This scenario is fictional and uses aliases, but stands on recent events, which enhances the likelihood of identification of actors. Fundamentally, the goal is to identify technical and governance failures rather than supporting specific political orientations or insert value-based judgement in the actors described.

LLM Usage Statement

LLMs by Anthropic, OpenAI and Mistral were first used to assess the plausibility of the scenario and estimate the number of deaths and material costs of the blackouts. Then, they provided feedback based on the evaluation criteria.