‘Regimes of Truth’ in Digital Public Diplomacy: How ChatGPT Structures Credibility, Legitimacy, and Conflict Narratives

Nilotpal Bhattacharjee
PhD Student, Department of Mass Communication, Assam University, Silchar

Abstract- Generative AI tools like ChatGPT are becoming part and parcel of the diplomatic process, but the role of these generative tools in building geopolitical knowledge has not been properly examined. Most of the current literature primarily frames AI as an instrument of balance-of-power, a governance dilemma, or a tool for disinformation amplification. However, this paper questions the role of ChatGPT's linguistic outputs in shaping political reality. Based on the Foucauldian Discourse Analysis (FDA) with references to Michel Foucault and Theo van Leeuwen, the paper uses ChatGPT (GPT-4o) outputs to perform a discourse analysis of three purposely selected case studies: the credibility ranking of state-funded international broadcasters, the comparative description of the Khalistan movement and ETA, and narration of the Israel-Palestine conflict. The prompts are run 5 times until they meet the threshold, generating 15 outputs from 3 structured prompts, in May 2025. The methodological design guaranteed the reproducibility and systematic coding in all cases. The analysis of three case studies highlighted recurring discursive patterns. In the first case study, ChatGPT showed hierarchical ordering of credibility by placing Western media outlets as more credible. In the second study, the chatbot exhibited classificatory anchoring, framing ETA as a terror group and Khalistan as a political movement. In the third case study, which discusses the Israel-Palestine conflict, ChatGPT adopted a neutral narrative, using objective, diplomatic language to describe the situation. These three patterns exhibited by ChatGPT do not reflect ideological intent but reproduce the historically sedimented regimes of truth embedded in the model's large and anonymous training corpus. This paper argues that ChatGPT is a producer of governmental discourse that silently structures diplomatic knowledge, especially when human involvement is not present. The paper also underlined the importance of AI literacy for international relations scholarship and foreign-policy practice, arguing that critical engagement with AI-mediated information production is essential.

I. INTRODUCTION

In May 1997, the US-based technology firm IBM’s “Deep Blue” supercomputer made history by defeating Russian chess grandmaster and the then world champion Garry Kasparov in a six-game match. This marked the first-ever incident in human civilisation when a computer beat a human in a game that requires players to use cognitive functions and intelligence. Little did people know that three decades later, artificial intelligence, such as generative chatbot ChatGPT, would be used by diplomats to learn about world events, draft engaging social media posts, quickly provide consular assistance, draft press statements, and analyse vast datasets necessary for negotiations (Kilburg, 2025; Manor, 2023). A 2019 Diplo Foundation report stated that AI is not only a “topic for diplomacy”, but also “a tool for diplomacy”, noting that “additional insights generated through AI applications can contribute to better foreign policy decision-making” (Diplo Foundation, 2019).

Previous studies in international relations framed AI in four categories: “balance of power, governance, disinformation, and ethics” (Bode, 2024). Several past researchers discussed the possibility of using AI by the military establishments of a state during a conflict (Payne, 2021) and in maintaining the “stability of the balance of power, international institutional order, and international norms” (Horowitz et al., 2024), while some highlighted AI governance, stressing the need for algorithmic accountability by global powers (Maas, 2023; Roberts et al., 2024). Certain studies raised concerns about the AI’s “cognitive” biases (Dubois et al., 2025; Koo et al., 2024; Rastogi et al., 2022), while some discussed amplification of AI-generated disinformation and misinformation (Wardle & Derakshan, 2017). Manor (2023) opined that ChatGPT may reinforce “stereotypes” and contribute to the ongoing “inequalities between the Global North and the Global South”. Caliskan (2023) mentioned that human-produced texts integrated into an AI system contain inherent biases that influence how an AI model generates text, images, or responses. So the majority of these studies shared an underlying assumption that AI systems are instruments or actors operating within a political field, whether as tools for balancing power structure, or governance mechanisms, or misinformation vectors. What is less examined is how AI-generated language itself participates in structuring that field. This paper seeks to analyse the “language” that ChatGPT generates in response to certain purposefully-selected prompts on geopolitical issues. The paper seeks to examine whether ChatGPT’s outputs governmentalise discourse on international issues through three discursive mechanisms, namely credibility hierarchies, legitimacy classifications, and narrative neutralisation, through three purposefully selected case studies.

This paper relies on French philosopher Michel Foucault's theories of discourse and governmentality to understand if ChatGPT’s presumed “neutral” outputs establish “regimes of truth” through a nuanced ordering mechanism. Foucault argued that knowledge and power are not separate, reigning domains (Foucault, 1977; 1980). The standards for credibility, legitimacy and responsibility don’t just spring up from nowhere. The institutional repetition, professional norms, and historically sedimented hierarchies stabilise this. The stabilisations are described by Foucault as what he calls “regimes of truth”; in which certain statements acquire authority because they circulate in order to be recognised as valid (Foucault, 1980).

Every LLM is trained on a massive, diverse corpus of text and data, and thus, when it generates answers to prompts, those answers are not truly its own. Although it is called artificial intelligence, generative AI cannot think as a human does, and its responses are based on the built-in knowledge it has. It simply operates through an archive of text shaped by previous power relations. Investigations into algorithmic bias usually zero in on quantifiable gaps in output, such as its political tilt, stereotyped depictions of particular groups, and partisan skew (Motoki, Neto, & Rodrigues, 2024). Understanding the biases is important, as generative AI may perpetuate and exacerbate existing inequalities. But, there is also a need to look beyond these measurable biases to analyse stronger structural effects. Even when an AI-generated text looks neutral or fair on the surface, it can still quietly repeat dominant ways of classifying and judging people, movements, and institutions.

This paper, therefore, asks a different set of questions. When ChatGPT is asked to rank state-funded international broadcasters by credibility, how does it decide what “credible” means? When it is asked to describe a separatist movement, how does it decide which actors seem legitimate and which do not? When it tells the story of a major, highly visible conflict, how does it assign or shift responsibility and blame? These are not just questions about whether the facts are correct. These are questions of discursive ordering.

Digital public diplomacy has raised the question of narrative competition, which is strategic communication and the mediation of national image (Melissen, 2005; Manor, 2019). As AI becomes more responsible for accessing background knowledge, rather than commercial news organisations, have helped constitute the discursive ecology that creates the interpretive terrain of public diplomacy. A ranked list, definitional label, and balanced conflict summary can serve as micro-briefings that set the frame for forthcoming official diplomatic messaging and contesting. The analysis of ChatGPT’s outputs using a Foucauldian Discourse Analysis (FDA) sheds light on how the outputs produced by ChatGPT are discursively engaged in the production of regimes of truth. To conduct this analysis, three case studies have been selected purposefully. The first is the credibility ranking of international broadcasters. The second is the comparative description of the Khalistan movement and the ETA. The last is the narration of the Israel-Palestine conflict. The cases gathered show a series of biases which have been given a hierarchical ordering, a classificatory anchoring, and a narrative neutralisation. However, there is a need to see these ‘biases’ more as augmentations of digitally mediated geopolitical knowledge, structured normalisation. As these systems become embedded in everyday informational routines, the quiet organisation of credibility, legitimacy, and responsibility, deserve closer scrutiny, not because the model shouts ideology, but because it rarely needs to.

II. THEORETICAL FRAMEWORK

This study employs Michel Foucault's interconnected theories of discourse and power/knowledge to interrogate generative AI outputs as constitutive forces in geopolitical meaning-making, rather than passive reflections of pre-existing realities. Foucault's radical departure from conventional epistemology lies in his rejection of any foundational divide between knowledge production and power relations, what he terms the “repressive hypothesis” that treats power as an external constraint on truth (Foucault, 1978). Instead, truth emerges relationally through historically contingent “regimes of truth”: procedural ensembles comprising “the ensemble of rules according to which the true and the false are separated and specific effects of power attached to the true” (Foucault, 1980, p. 131). These regimes operate via institutional repetition, validation mechanisms, and discursive circulation that render certain statements recognisable as authoritative, while marginalising others as implausible or illegitimate.

Power can also be productive, facilitating new ways of thinking, acting, and living. Knowledge is the foundation of power. In “*Discipline and Punish: The Birth of the Prison”,* Foucault (1977) argued that the people in power do not use or promote knowledge just because it is “useful”; instead, “power produces knowledge”, and they depend on each other to exist. Knowledge generates power, and power, in turn, produces knowledge in a reciprocal relationship. People with power use knowledge to control or increase their dominance over others, while power also “shapes the conditions for creating and deploying knowledge” (Foucault, 1980). Power reinforces its influence per its “depersonalised interests and intentions” by shaping the conditions of knowledge production and validation, which may include, among others, “social structures of universities, scientific publishing norms, government funding, intellectual property laws,” and even colonial practices that exploit both people and ecosystems as resources for technology development (Burrell & Metcalf, 2024). So, knowledge is utilised as a means of control.

In international relations theory, Foucault's ideas challenge positivist and rationalist paradigms, realism's material power distributions, liberalism's institutional bargains, and constructivism's intersubjective norms by repositioning discourse as the ontological ground of political possibility (Edkins, 1999; Milliken, 1999; Weber, 2010). Language does not merely describe or mediate interests but constitutes the conditions of intelligibility: sovereign subjects, legitimate interventions, credible threats emerge as effects of discursive rules governing what can be said, by whom, and with what effects (Foucault, 1972; Walker, 1993; Campbell, 1992). This constitutive function manifests archaeologically, as Foucault outlines in The Archaeology of Knowledge, discourses function as rule-bound archives that delimit “the totality of relations that can be discovered, for a given period, among the elements of a discursive formation” (Foucault, 1972, p. 38), pre-structuring geopolitical fields before strategic action commences.

Foucault's later lectures introduce governmentality as the analytical bridge between micro-discourses and macro-political order: “the conduct of conduct,” or the dispersed arts of population management through techniques of normalisation, categorisation, and measurement rather than juridical sovereignty (Foucault, 2004; Lemke, 2001). Governmentality proliferates through discourses that render complex phenomena calculable, grids of specification, procedures of hierarchization, and mechanisms of control, transforming political contestation into administrable domains (Foucault, 2007). In contemporary IR, this illuminates how security rationalities, development discourses, and humanitarian norms govern global conduct by framing problems in “manageable” terms, such as threat rankings, vulnerability indices, and intervention thresholds (Dillon, 1995; Rasmussen, 2004).

This theoretical triad, such as “regimes of truth”, constitutive discourse, and governmentality, directly addresses a critical lacuna in AI-international relations scholarship. Predominant studies adopt instrumental framings, conceptualising AI as exogenous technology acting within assumed political fields: balance-of-power multipliers (e.g., autonomous weapons), governance dilemmas (e.g., algorithmic accountability), disinformation amplifiers, or ethical quandaries (Bode, 2024; Horowitz, 2018; Maas, 2019). Such approaches presuppose stable discursive preconditions, neutral categories of “credibility,” “legitimacy,” “responsibility”, while neglecting how AI language itself produces these coordinates (Cavelty, 2020; Egloff, 2021).

FDA treats generative AI outputs as discursive events that stabilise truth regimes through specific operations: (1) hierarchical ordering, converting contested qualities into ordinal scales; (2) classificatory anchoring, where primary labels pre-structure interpretive frames; and (3) narrative normalisation, diffusing agency through symmetrical or passive constructions (van Leeuwen, 2008; Keller et al., 2008). Unlike quantitative bias audits measuring partisan skew against ground-truth baselines, FDA excavates structural effects: how training on power-laden corpora recombines dominant discourses, rendering geopolitical asymmetries technical and inevitable (Chandler, 2018). Recent IR applications confirm the FDA's purchase of digital texts, from cyber-security rationalities to climate governance discourses (Pouliot, 2015; Gronau & Schmidtke, 2016).

This framework positions ChatGPT not as a neutral aggregator but as a governmental discourse-producer, embedded within and reproducing international informational hierarchies (Thussu, 2020). By operationalising the FDA across three case studies, the analysis reveals AI's “quiet organisation” of diplomatic knowledge beyond overt ideology, through the mundane work of lists, labels, and summaries.

III. METHODOLOGY

The current work was based on a qualitative interpretive paradigm, Foucaultian Discourse Analysis (FDA), to challenge the linguistic production of a geopolitical reality. In contrast to the general literature on AI that focuses on accuracy or bias, our question sought to explain why language itself can create the perceived plausibility, authentic personhood, and attribution of blame (Foucault, 1972). This area of interest is relevant because the use of AI-generated briefings by diplomatic practitioners has become increasingly prevalent, and framing effects can become a reality even before observers are aware of them.

ChatGPT outputs to certain pre-defined purposefully made prompts were collected during the last two weeks of May 2025 via the general interface of the ChatGPT (GPT-4o) chatbot. Prompts to custom, system messages, or parameter adjustments were also avoided to record the desired canonical order of interaction that a foreign-service officer may undergo when preparing for a crisis late at night. By choosing three purposive cases, four different mechanisms of discursive were charted out:

Credibility Pyramids: Ranking of state-funded international broadcasters by credibility.

Type of Legitimacy: A comparison of the Khalistan movement (India) and the ETA (Spain).

Narrative Neutralisation: Understanding ChatGPT’s version of the Israel-Palestine conflict.

I ran the prompts five times. The fourth and fifth analyses reproduced the same hierarchies, classifications, and framing patterns, indicating recurrence in the discourse (see Keller, 2006; Bohnsack, 2008, on reconstructing recurrent structural features in discourse), and thus it was not required to run the prompts again. The obtained corpus consisted of 15 outputs (three prompts 5 runs).

IV. ANALYTICAL PROCEDURES

The analysis was carried out in three intersecting stages rather than in a linear fashion.

1. This analysis of textual structuring analysed formal discursive practices such as hierarchy of lists, initial classificatory terms, passive forms, and reciprocal syntactic structures, which create power distributions.

2. The discursive reproduction analysis was used to establish institutional common-sense markers that have been appropriated into existing repertoires, including credibility validation practices (widely respected), security classification norms (banned terrorist organization), as well as diplomatic neutrality conventions (both sides).

3. The social embedding analysis placed the revealed patterns in the larger geopolitical knowledge hierarchies, which follow the alignment with the

standards of Western media dominance, Global North validation, and humanitarian framing conventions.

Coding was done using NVivo 14 qualitative analysis software to do systematic memoing and pattern tracking. The inter-rater reliability test was done with a discourse analysis colleague on a 25 per cent random sample, which resulted in 91 per cent agreement on the main discursive operations.

Positionality and management of bias:

The greatest guiding principle was transparency. Being the researchers operating in India at that time of heightened India-Canada Khalistan tensions, our positionality heightened our sensitivity to Western media framing and posed a risk of confirmation bias. There were mitigation strategies such as:

(a) The same prompts were used systematically across all cases,

(b) The results were compared with both Western and non-Western reference points,

(c) The researcher actively looked for counterexamples where Khalistan and ETA might be treated in a genuinely similar way, and

(d) Analytic notes and interpretations were reviewed by an external reader.

Limitations:

The study was dependent on the outputs of GPT-4o in May 2025 to certain purposefully curated prompts. The model update, variation on multilingualism, and training data not owned by the partner are not investigated. The monolingualism in use can be a constraint when generalising to other language environments, such as Mandarin or Arabic. Inaccessibility to training data limits the depth of analysis. Triangulation helps to eliminate subjectivity, although the FDA already adheres to interpretive interaction. The repetition of five runs per prompt creates evidential, not statistical, generalisation. Overall, this approach to methodology helped uncover discursive processes that underpin geopolitical governmentalisation: credibility rating, legitimacy labelling, and conflict neutralisation can occur well in advance of diplomatic players formally entering the decision-making process.

V. FINDINGS

Case Study 1: Credibility Rankings and Hierarchy Ordering

When asked to rank the most credible state-funded news sources worldwide, ChatGPT rated Germany's Deutsche Welle (DW) and Voice of America (VOA) as the most credible. Other options included France 24, Russia Today, Doordarshan (India's public broadcaster) and Al Jazeera (funded by the Qatari government). The prompt was: Please rank the following news sources based on their credibility: Deutsche Welle, Voice of America, France 24, Russia Today, Doordarshan, and Al Jazeera. Given that ChatGPT provides varying answers to the same question when requested consecutively, the researchers subsequently submitted the same prompt. The responses from ChatGPT to the same prompt submitted separately were almost identical, except that Russia Today was ranked higher than DD News on the first occasion, while the positions were reversed the second time.

DW (Deutsche Welle), VOA (Voice of America) and France 24 are often referred to as independent, publicly funded, and have a good reputation for reliable reporting, and are considered respected international broadcasters. To a large extent, these outlets position themselves as neutral ones, with only faint or situation-specific disclaimers, like an American or slight French and European bias to their coverage. As a result, they are placed as the default, reliable providers of European and international news or foreign events. In comparison, DD News and Russia Today (RT) start with a clear focus on government ownership, state interests, and claims of false information and disinformation. DD News has been recognised as credible in its factual reporting on India, but in quick succession, it is cut short by fears that the news outlet lacks full independence in reporting sensitive domestic political affairs. The RT is largely presented in the light of its bias towards the Russian government.

These characterisations indicate a hierarchical positioning in which Western or Western-aligned media houses are framed as factual and neutral, whereas non-Western media are defined mostly as biased towards the state governments. The output aligns with Foucault's regime of truth, in which credibility is not decided by open and fair debate but by a pre-sedimented hierarchy, where certain outlets are positioned as authoritative through repetitive validation. In simple terms, a few top media outlets are automatically trusted because they have built reputations over time, and these "authoritative" sources maintain their high status simply by repeating and validating one another's information, creating a "hierarchy". The ranking (table 1) enacts governmentality by attaching undefined descriptors such as "independent," "balanced," and "professional" to the upper tiers, while "state-aligned" and "editorial constraints" contextualise the lower ones (Foucault, 2004). No state-owned broadcaster receives outright criticism for its biases towards its owners, yet the vertical arrangement performs a form of differentiation: superior ranks are given unmerited legitimacy, while lower ones have to be explained politically.

This hierarchy shows its compatibility with Global North dominance trends through external benchmarking. The 2025 Press Gazette ranks the BBC (1.2 billion monthly visits) and CNN as the best news outlets, and the SCImago media reputation index places The Guardian and Reuters as trusted news sources over non-Western websites. The 2025 trust survey by YouGov also labels the BBC/PBS as the most trustworthy, followed by state-funded ones, such as RT, which is less trusted by Western audiences. Therefore, the above credibility ranking by ChatGPT is not something the chatbot has made on its own. It has only recombined digitised reputational common-sense, in which Western public-service models ("charter-mandated objectivity") signify credibility, while others are disqualified as "government funded" (Thussu, 2020).

This effect is magnified by the procedural tone of the response discursively. While describing these outlets, ChatGPT does not mention any qualifiers, such as "perceptions may vary" and "methodology includes certain parameters." Instead, evaluative language has been used, such as "rigorous" and "respected" for higher-ranked outlets, while lower-ranked ones were tagged as "despite strengths, limited by...". Van Leeuwen's legitimation analysis identifies this as "authorisation by expert validation" versus "qualification by circumstance," a classic hierarchical technique (van Leeuwen, 2008). Van Leeuwen's four categories of legitimation included: "authorisation", which is a "reference to authority figures, custom or law"; "moral evaluation", which is a "reference to value systems"; "rationalization", which is a "reference to the utility or purpose of particular actions and to social knowledges that give them cognitive validity"; and "mythopoesis" which is a "reference to narratives that reward legitimate actions" (Harjuniemi, 2021).

ChatGPT's answers to the prompts establish a ranking in which credibility becomes technical, quantifiable, and gradeable, rather than geopolitically contentious. This issue is pertinent to the discourse of digital public diplomacy. Documenting the increasing use of AI in narrative competition, Bjola and Manor (2023) introduce an aspect of micro-briefings that establishes framing parameters of official messages. ChatGPT's rankings not just summarise reputation but turn it into some form of diplomatic shorthand, placing Deutsche Welle as a credible ground rule and RT/DD News as political extremes. This state of affairs makes the informational dominance of the West administratively rational and has no justification except its place in the list. This trend continues through prompt repetitions: wording (high journalistic standards and strong fact-checking) changes, but the hierarchy (Deutsche Welle/Voice of America at the top, RT/DD News at the bottom) does not. This repetition proves discursive stability beyond stochastic generation, entrenching already existing power geometries in the processes by which AI creates truth.

Case Study 2: Khalistan vs. ETA Classification

The second inquiry entailed asking ChatGPT to explain the Khalistan movement (India) and ETA (Spain). ETA is immediately introduced as a "Basque separatist group… responsible for over 800 deaths" and "widely considered a terrorist organization," with emphasis on "bombings, assassinations, and kidnappings" and its status as "banned." The Khalistan movement, by contrast, is framed as "a political and separatist movement… calling for the creation of an independent Sikh state," grounded in "historical grievances," "the trauma of Partition," "Operation Blue Star," and "subsequent anti-Sikh riots." The first categorisation predetermined the subsequent discourse, in which ETA was categorised in a security framework and Khalistan in a political framework.

If it is seen through Foucault's lens, this dichotomy amounts to a differentiated positioning of subjects within a discursive field: "terrorist organization" inserts ETA into a grid where neutralisation and security management appear self-evident, while "political and separatist movement" locates Khalistan within a repertoire of negotiation, grievance and representation. Van Leeuwen's legitimation framework helps specify how this operates at the level of justification. ETA is legitimised primarily through authority and legality, using terms such as "banned," "widely considered a terrorist organisation," and the focus on casualties aligns with what van Leeuwen describes as authorisation and moral evaluation, supporting suppression as the normal response. Khalistan draws on moral and historical legitimation: references to "historical grievances," "partition", "Operation Blue Star", and "anti-Sikh violence" invoke a narrative of suffering and injustice that affords the movement a degree of interpretive sympathy. Grammatical agency also diverges: ETA typically appears as an active subject carrying out bombings and assassinations, whereas Khalistan is more often the context for actions ("debates," "demands," "grievances"), with agency distributed to states and diaspora actors.

This trend is supported by empirical research in media representation. Studies of large language models and extremism show that these systems inherit strong associations between certain religious identities, especially "Muslim" and terrorism, and more broadly reproduce existing Western security framings in how they describe political violence (Abid et al., 2021; Dong et al., 2024). A 2023 analysis of global coverage of the Khalistan movement revealed that recent reporting increasingly frames Khalistan as a form of Sikh diaspora activism (Singh, 2023). Across the repeated responses, ChatGPT consistently characterised ETA as a terrorist, whilst Khalistan was always framed within a political context. The consequences of diplomatic interactions based on a chatbot's output could be far-reaching.

Findings: Case Study 3: Israel-Palestine Neutralisation

When asked to describe the Israel-Palestine conflict, ChatGPT used textbook diplomatic phrasing. Formulations such as "long-standing conflict" over rival national claims, mutual displacement and violence, and the recidivism of successive retaliation were recurring patterns in the model's responses and reflected the language regularly found in United Nations briefings.

It is interesting to note that the model's first point is that it flattened massive asymmetries. This flattening is typified by the confusion of the Nakba in 1948 with the displacement of 750,000 Palestinians into a generic statement of what occurs during the creation of Israel, and a parallel statement of the Jews fleeing persecution. Equally, the systematic juxtaposition between the number of casualties during the October 7th ("1,200 Israeli dead" and "over 40,000 Palestinian dead"), and Gaza operations (over 40,000 Palestinian dead) with the unbiased recognition of the fact that Hamas initiated attacks and Israel made military strikes, manages to do away with references to the 17-year blockade, settlement numbers (more than 700,000), and ratio of casualties which are stacked heavily in favour of the Israeli side. The argument always leads to an account of civilian anguish on either side.

This is what Foucault (2004) refers to as governmentality, whereby power decentralises responsibility through language, thereby making the occupation a simple argument about territory and settlements. The comparison of rockets and airstrikes as the same activity, as well as the insistence on "both sides" framing, shields specific responsibility and establishes a cycle of the peace process that could not be achieved due to the avoidance of assigning individual accountability.

This observation is supported by an analysis of institutional data. Scholars have noted the Western media's bias towards Israel in its coverage of the conflict (Al-Najjar & Zaid, 2025). Ghani (2025) noted that the Western media outlets such as "CNN, BBC, The New York Times, and other major outlets" repeatedly provided "one-sided coverage, selective terminology, and lack of balanced representation" of the Israel-Palestine conflict.

It is also not easy to discuss the conflict through a legal lens because whether to frame it as an international or non-international armed conflict depends on which legal framework is applied (Malik, 2023; Shereshevsky, 2024; Steenberghe, 2025). There have been a lot of discussions over the blockade of Gaza and the alleged expansion of settlements in the West Bank from a policy perspective, but it has not been examined as legal violations. The multilateral forum, especially the United Nations, emphasised the "both parties'" participation to halt the conflict and violence and to adhere to humanitarian law (United Nations, 2024), even though one side has been the major victim. LLMs are trained on this kind of material. In their study on the LLM evaluation of the coverage of the Israel-Hamas conflict by BBC and The Guardian, Chandra et al., (2026) found "larger political bias shifts" and noted that "LLMs are shaped not only by their training data and architecture, but by underlying worldviews with associated political biases". It is done in conflicts just to frame it from a neutral perspective and avoid assigning clear responsibility. These tendencies are not intentional decisions made by the chatbot. Instead, they come from the training data, the same media coverage, legal commentary, and diplomatic texts where asymmetry has often been hidden. The bias is built into the structure itself.

A consistent structural pattern emerged through repeated testing of the two prompts. The conflict was framed as (1) a source of rival national claims, (2) a vicious cycle of violence, and (3) blocked negotiations. Minor lexical variations, such as replacing 'intractable' with 'deeply rooted,' did not alter this symmetrical framework. The stable results suggest that the idea of neutrality is not superficial but a fundamental aspect of how the model encodes knowledge.

VI. DISCUSSION

All three case studies cited above indicate that ChatGPT is not only extracting or summarising information on international affairs, but also actively involved in the conceptualisation of political reality. In each case, ChatGPT's outputs appear to be structured epistemic arrangements that are in close correspondence with Foucault's concept of regime of truth. The outputs are based on historical archives that determine what can be articulated, what can be made intelligible, and how power and responsibility are allocated in a body of knowledge.

The credibility ranking in the first case study is not established by a systematic comparison of editorial norms. It was informed by a hierarchical structure that makes Western or West-oriented broadcasters appear neutral and objective, and views Global South media outlets through the prism of state bias and propaganda. Deutsche Welle and VOA are often described as independent and reliable, whereas DD News and RT are largely characterised by their association with state authorities. This stratification is not the result of a neutral algorithm, but rather re-conceptualises language already saturated with established geopolitical relations of power. Foucault argued that what we take to be "self-evident" or "true" is actually produced by power and historical context, rather than being inherently obvious or universal. ChatGPT's credibility rankings exemplify this when it presents a politically influenced hierarchy as a technical outcome.

Case Study 2 shows how discourse controls which actors assume specific roles in international affairs. The first labels, "terrorist organization" for ETA and "political and separatist movement" for Khalistan, serve as formation rules. Those rules limit later statements and create distinct interpretive grids. This mechanism is explained by Van Leeuwen's analytical framework: ETA gains legitimacy through formal prohibition and its established casualty statistics, whereas Khalistan gains moral legitimacy through the expression of its historical grievances and support from the diaspora. Those roles are not neutral accounts - they determine whether an actor counts as a political partner or faces dismissal as a security threat. AI classification, therefore, becomes a site where governance operates in silence before human judgment appears.

Case Study 3 demonstrates that the Israel-Palestine conflict is treated as a technical issue of balancing two equal claims instead of a situation shaped by built in imbalance. Each ChatGPT reply repeats the same three-step story: rival historical claims, mutual rounds of violence, and frozen peace talks. This pattern, which reappears with only small word changes after many different prompts, matches a routine already noted in Western press reports, UN Security Council texts and legal debates - those sources also favour balanced language even when power plus control are clearly unequal. From Foucault's perspective, ChatGPT restates the standard store of international mediation knowledge: the clash becomes a problem of negotiation between apparently matched sides, while facts such as military occupation, blockade, and unequal sovereignty become harder to state or contest. In every case studied, the same governing effect appears: the AI sets the map in advance; it marks who counts as trustworthy, who holds legitimacy, which violence is labelled as terror, and which is labelled as protest, before any human expert, envoy, or student begins work. Foucault's approach reveals that power acts not only through open force or stated belief but through the silent reuse of categories, ranks, and also story patterns.

VII. CONCLUSION

This study aimed to find out how ChatGPT builds a picture of world politics in three areas: how trustworthy broadcasters appear, how separatist movements are labelled, and how major conflicts are presented. It used Foucauldian Discourse Analysis as the main method and also applied van Leeuwen's legitimation framework as a secondary tool. The study reveals that ChatGPT's responses are not free of political bias. Each response is a piece of discourse that repeats plus reinforces specific versions of truth drawn from the data on which the model was trained.

The results matter for both research and policy - for academics in international relations and media studies, the study shows that texts produced by AI should be examined with the same care given to government papers, press statements or diplomatic notes. The main danger is not that the text contains false facts, since such errors can be spotted and fixed. The bigger risk lies in what Foucault calls the shaping of what becomes speakable - the silent removal of other ways to judge credibility, legitimacy or conflict. When “complex history” takes the place of a structural explanation, when “both parties” stands in for a clearly identified occupying power, when “political movement” and “terrorist organisation” place parties into entirely separate governance categories, those are not harmless shortcuts. They are political choices that arrive disguised as neutral information. For professionals, those results matter, as generative AI tools are increasingly used in intelligence reports, diplomatic messages, and education, and the ways those tools describe the world risk becoming standard. Analysts who use AI-generated summaries without questioning the categories and assumptions built into them may find that their range of political options has already been limited before they make any decisions. This shows that media literacy and skills in discourse analysis are not just academic interests. They are essential for anyone working with AI in international relations.

Future research should expand this work - studies that compare different AI models - like Claude, Gemini and systems developed in China, Russia, besides India - would help determine whether the patterns found in ChatGPT are unique to it or appear across other systems.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my supervisor, Prof. Paromita Das, for her guidance, encouragement, and insightful feedback. I also extend my thanks to the two independent coders who assisted with the coding process and helped ensure the analysis's reliability.

ETHICAL APPROVAL

This study did not require formal ethical approval as it involved no human participants, personal data, or sensitive information.

REFERENCES

[1] Abid, A., Farooqi, M., & Zou, J. (2021). Persistent Anti-Muslim Bias in Large Language Models. Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, 298–306. https://doi.org/10.1145/3461702.3462624

[2] Al-Najjar, A., & Zaid, B. (2025). Western media's ethical collapse: Silencing Gaza's voice. Third World Quarterly, 1–20. https://doi.org/10.1080/01436597.2025.2552361

[3] Bjola, C., & Manor, I. (2023). ChatGPT: The end of diplomacy as we know it. Global Policy Journal.https://www.globalpolicyjournal.com/blog/25/04/2023/chatgpt-end-diplomacy-we-know-it

[4] Bode, I. (2024). AI Technologies and International Relations: Do We Need New Analytical Frameworks? The RUSI Journal, 169(5), 66–74. https://doi.org/10.1080/03071847.2024.2392394

[5] Bohnsack, R. (2008). Rekonstruktive Sozialforschung: Einführung in qualitative Methoden (7th ed.). Barbara Budrich.

[6] Burrell, J., & Metcalf, J. (2024). Introduction for the special issue of “Ideologies of AI and the consolidation of power”: Naming power. First Monday. https://doi.org/10.5210/fm.v29i4.13643

[7] Caliskan, A. (2023). Artificial Intelligence, Bias, and Ethics. Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, 7007–7013. https://doi.org/10.24963/ijcai.2023/799

[8] Campbell, D. (1992). Writing security: United States foreign policy and the politics of identity. University of Minnesota Press.

[9] Cavelty, M. D. (2020). Cyber-security. In A. Collins (Ed.), Contemporary security studies (5th ed., pp. 291–307). Oxford University Press.

[10] Chandler, D. (2018). Ontopolitics in the Anthropocene: An introduction to mapping, sensing and hacking. Routledge.

[11] Chandra, R., Chen, H., Zhang, Y., Chen, J., & Wu, Y. (2026). An evaluation of LLMs for political bias in Western media: Israel-Hamas and Ukraine-Russia wars (arXiv:2601.06132). arXiv. https://doi.org/10.48550/arXiv.2601.06132

[12] Dillon, M. (1995). Sovereignty and governmentality: From the problematics of the “new world order” to the ethical problematic of the world order. Alternatives: Global, Local, Political, 20(3), 323–368. https://doi.org/10.1177/030437549502000303

[13] Diplo Foundation. (2019). Mapping the challenges and opportunities of artificial intelligence for the conduct of diplomacy. https://www.diplomacy.edu/resource/mapping-the-challenges-and-opportunities-of-artificial-intelligence-for-the-conduct-of-diplomacy/

[14] Dong, B., Lee, J. R., Zhu, Z., & Srinivasan, B. (2024). Assessing Large Language Models for Online Extremism Research: Identification, Explanation, and New Knowledge (arXiv:2408.16749). arXiv. https://doi.org/10.48550/arXiv.2408.16749

[15] Dubois, Y., Galambosi, B., Liang, P., & Hashimoto, T. B. (2025). Length-Controlled AlpacaEval: A Simple Way to Debias Automatic Evaluators (arXiv:2404.04475). arXiv. https://doi.org/10.48550/arXiv.2404.04475

[16] Edkins, J. (1999). Poststructuralism and international relations: Bringing the political back in. Lynne Rienner Publishers.

[17] Egloff, F. J. (2021). Contested public attributions of cyber incidents and the role of states. Contemporary Security Policy, 42(1), 55–81. https://doi.org/10.1080/13523260.2020.1843571

[18] Foucault, M. (1972). The archaeology of knowledge. Pantheon Books.

[19] Foucault, M. (1977). Discipline and punish: The birth of the prison (1st ed.). Vintage Books.

[20] Foucault, M. (1978). The history of sexuality, Vol. 1: An introduction. Pantheon Books.

[21] Foucault, M. (1980). Power/knowledge: Selected interviews and other writings, 1972–1977 (C. Gordon, Ed.; C. Gordon, L. Marshall, J. Mepham, & K. Soper, Trans.). Pantheon Books.

[22] Foucault, M. (2004). Security, territory, population: Lectures at the Collège de France, 1977–1978. Palgrave Macmillan.

[23] Foucault, M. (2007). The birth of biopolitics: Lectures at the Collège de France, 1978–1979. Palgrave Macmillan.

[24] Ghani, F. (2025). The sharp contrast: How Israeli and Western media cover the war on Gaza. Al Jazeera Media Institute. https://institute.aljazeera.net/en/ajr/article/3081

[25] Gronau, J., & Schmidtke, H. (2016). The quest for legitimacy in world politics — international institutions' legitimation strategies. Review of International Studies, 42(3), 535–557. https://doi.org/10.1017/S0260210515000492

[26] Harjuniemi, T. (2021). The “hierarchy of credibility” among economic experts: Journalists' perceptions of experts with varying institutional affiliations. Journalism Practice, 16(8), 1635–1652. https://doi.org/10.1080/17512786.2021.1910985

[27] Horowitz, M. C. (2018). Artificial intelligence, international competition, and the balance of power. Texas National Security Review, 1(3), 36–57. https://doi.org/10.26153/tsw/865

[28] Horowitz, M. C., Pindyck, S., & Mahoney, C. (2024). AI, the international balance of power, and national security strategy. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), The Oxford handbook of AI governance (p. 0). Oxford University Press.

https://doi.org/10.1093/oxfordhb/9780197579329.013.55

[29] Keller, R. (2006). Wissenssoziologische Diskursanalyse. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 7(1), Art. 11. https://doi.org/10.17169/fqs-7.1.75

[30] Keller, R., Hirseland, A., Schneider, W., & Viehöver, W. (Eds.). (2008). Handbuch Sozialwissenschaftliche Diskursanalyse (2nd ed.). VS Verlag für Sozialwissenschaften.

[31] Kilburg, D. (2025). AI use cases for diplomats: Applying artificial intelligence to diplomacy (1st ed.). Chapman & Hall. https://www.routledge.com/AI-Use-Cases-for-Diplomats-Applying-Artificial-Intelligence-to-Diplomacy/Kilburg/p/book/9781041008354

[32] Koo, R., Lee, M., Raheja, V., Park, J. I., Kim, Z. M., & Kang, D. (2024). Benchmarking cognitive biases in large language models as evaluators. In L.-W. Ku, A. Martins, & V. Srikumar (Eds.), Findings of the Association for Computational Linguistics: ACL 2024 (pp. 517–545). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-acl.29

[33] Lemke, T. (2001). "The birth of bio-politics": Michel Foucault's lecture at the Collège de France on neo-liberal governmentality. Economy and Society, 30(2), 190–207. https://doi.org/10.1080/03085140120042271

[34] Maas, M. (2019). How viable is international arms control for military artificial intelligence? Three lessons from nuclear weapons. Contemporary Security Policy, 40(3), 285–311. https://doi.org/10.1080/13523260.2019.1576464

[35] Maas, M. (2023, November 1). Advanced AI governance: A literature review of problems, options, and proposals. Institute for Law & AI. https://law-ai.org/advanced-ai-gov-litrev/

[36] Malik. (2023, November 24). Classification of the Israel-Palestine conflict under the laws of war. Opinio Juris. https://opiniojuris.org/2023/11/24/classification-of-the-israel-palestine-conflict-under-the-laws-of-war/

[37] Manor, I. (2019). The digitalization of public diplomacy. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-04405-3

[38] Manor, I. (2023, April 7). ChatGPT and the threat to diplomacy. E-International Relations. https://www.e-ir.info/2023/04/07/opinion-chatgpt-and-the-threat-to-diplomacy/

[39] Melissen, J. (Ed.). (2005). *The new public diplomacy: Soft power in international relations*. Palgrave Macmillan.

[40] Milliken, J. (1999). The study of discourse in international relations: A critique of research and methods. European Journal of International Relations, 5(2), 225–254. https://doi.org/10.1177/1354066199005002003

[41] Motoki, F., Pinho Neto, V., & Rodrigues, V. (2024). More human than human: Measuring ChatGPT political bias. Public Choice, 198(1), 3–23. https://doi.org/10.1007/s11127-023-01097-2

[42] Payne, K. (2021). I, Warbot: The dawn of artificially intelligent conflict. Oxford University Press.https://doi.org/10.1093/oso/9780197611692.001.0001

[43] Pouliot, V. (2015). Practice theory and the study of diplomacy: A research agenda. Cooperation and Conflict, 51(1), 3–22. https://doi.org/10.1177/0010836715574913

[44] Rasmussen, M. V. (2004). "It sounds like a riddle": Security studies, the war on terror and risk. Millennium: Journal of International Studies, 33(2), 381–395. https://doi.org/10.1177/03058298040330020601

[45] Rastogi, C., Zhang, Y., Wei, D., Varshney, K. R., Dhurandhar, A., & Tomsett, R. (2022). Deciding fast and slow: The role of cognitive biases in AI-assisted decision-making. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW1), 1–22. https://doi.org/10.1145/3512930

[46] Roberts, H., Hine, E., Taddeo, M., & Floridi, L. (2024). Global AI governance: Barriers and pathways forward. International Affairs, 100(3), 1275–1286. https://doi.org/10.1093/ia/iiae073

[47] Shereshevsky, Y. (2024, June 18). Armed conflict classification in the ICC prosecutor's request for arrest warrants. Just Security. https://www.justsecurity.org/96914/armed-conflict-classification-icc-palestine-situation-gaza/

[48] Singh, N. (2023, September 29). 'Diaspora illusion' that's 'helping Modi': How global media views the Khalistan movement. Newslaundry.

https://www.newslaundry.com/2023/09/29/diaspora-illusion-thats-helping-modi-how-global-media-views-the-khalistan-movement

[49] Steenberghe, R. van. (2025, December 2). The ICJ obligations of Israel advisory opinion — qualifying Israel as an occupying power in the Gaza Strip. Lieber Institute West Point. https://lieber.westpoint.edu/qualifying-israel-occupying-power-gaza-strip/

[50] Thussu, D. K. (2020). *International communication: Continuity and change* (3rd ed.). Bloomsbury Academic.

[51] United Nations. (2024). *Adopting Resolution 2735 (2024): Security Council welcomes new Gaza ceasefire proposal*. https://press.un.org/en/2024/sc15723.doc.htm

[52] van Leeuwen, T. (2008). *Discourse and practice: New tools for critical discourse analysis*. Oxford University Press.

[53] Walker, R. B. J. (1993). *Inside/outside: International relations as political theory*. Cambridge University Press.

[54] Wardle, C., & Derakshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe. https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html

[55] Weber, C. (2010). *International relations theory: A critical introduction* (3rd ed.). Routledge.