

Faculty of Political Science
University of Belgrade
Master in Peace, Security and Development



Master's Thesis

**Epistemological Security in the Age of AI: The Case of
Deepfakes in Serbian Politics**

Mentor:

Prof. Filip Ejodus, PhD

Candidate:

Emilija Mladenović, 3011/2023

Belgrade, 2026

Table of Contents

Introduction	6
1. Literature Review	9
1.1 AI in IR: From Military Applications to Information Environments	9
1.2 AI, Political Communication and Democratic Trust	11
1.3 Deepfakes as a Political and Democratic Challenge	13
1.4 Hybrid Regimes, Digital Authoritarianism and Media Capture	18
1.4.1 Serbia: Democratic Backsliding and Media Capture	20
1.5 Research Gap and Contribution	23
2. Theoretical Framework	25
2.1 Ontological Security as a Conceptual Starting Point	25
2.2 Epistemological Security: Definition and Conceptual Boundaries	26
2.2.1 Distinguishing Epistemological Security from Adjacent Concepts	29
2.3 Level of Analysis	32
2.4 Deepfake-Related Truth Subversion: Offensive Fabrication and Defensive Dismissal	34
2.4.1 Truth, Interpretation and Defactualization	34
2.4.2 Truth-Subversion Practices	36
2.5 Epistemic Authority, Verification Networks and Institutional Conditions	38
2.6 Operationalisation: Four Analytical Dimensions	42
3. Methodology	45
3.1 Research Design	45
3.2 Data Collection	46
3.3 Analytical Approach	48
3.4 Methodological Limitations	51
4. Empirical Analysis: Deepfake-Related Truth Subversion in Serbian Politics	52
4.1 Political, Media and Institutional Context	52
4.2 Offensive Fabrication Practices (Dimension 1)	55
4.3 Defensive Dismissal Practices (Dimension 2)	59
4.4 Verification, Institutional Response and Public Contestation (Dimensions 3 and 4)	60
5. Discussion	65
5.1 Offensive Fabrication and Defensive Dismissal as Connected Practices	65
5.2 Credible Verification, Institutional Uptake and Dissemination Power under Media Capture	68
5.3 Selective Epistemological Security	70

5.4 Broader Implications: Factual Truth, Accountability and AI	75
5.5 Limitations and Counterarguments	78
Conclusion	80
References	82

List of Abbreviations

AI	Artificial Intelligence
ANEM	Association of Independent Electronic Media
BIRN	Balkan Investigative Reporting Network
BIRODI	Bureau for Social Research
BTI	Bertelsmann Transformation Index
CEPA	Centre for European Policy Analysis
CRTA	Centre for Research, Transparency and Accountability
FDU	Faculty of Dramatic Arts
IR	International Relations
KRIK	Crime and Corruption Reporting Network
LIO	Liberal International Order
NUNS	Independent Journalists' Association of Serbia
ODIHR	Office for Democratic Institutions and Human Rights
OHCHR	Office of the United Nations High Commissioner for Human Rights
OSCE	Organization for Security and Co-operation in Europe
OSS	Ontological Security Studies
PACE	Parliamentary Assembly of the Council of Europe
REM	Regulatory Body for Electronic Media

RTS	Radio Television of Serbia		
SNS	Serbian Progressive Party		
SSP	Party of Freedom and Justice		
V-Dem	Varieties	of	Democracy

ABSTRACT

Research on political deepfakes has largely examined synthetic media as instruments of fabrication and the liar's dividend, through which authentic evidence can be dismissed as artificially generated. Less attention has been paid to how these offensive and defensive practices interact within political environments marked by media capture and compromised institutional autonomy. This thesis addresses that gap through Adler and Drieschova's concept of epistemological security, examining how AI-generated and AI-contested political evidence can challenge the shared procedures through which contested political facts are authenticated and authoritatively resolved. Using a qualitative single-case study, the thesis compares seven deepfake-related political episodes in Serbia between August 2023 and August 2025. The findings show that offensive fabrication and defensive dismissal can operate as connected components of a broader truth-subversion dynamic, a connection directly documented in one episode and treated as a plausibility claim across the wider chronology. Their significance for epistemological security depends not only on technological sophistication, but also on the relationship between credible verification, institutional uptake and dissemination power. This thesis conceptualises selective epistemological security as a specific institutional configuration of epistemological insecurity: the politically uneven activation of institutional procedures through which credible verification receives institutional uptake and authenticity disputes receive authoritative resolution.

Keywords: deepfakes; epistemological security; truth subversion; offensive fabrication; defensive dismissal; media capture; Serbia

Introduction

Deepfakes are most often understood as instruments of deception: tools for fabricating convincing but false content. Yet their political significance may lie as much in the opposite direction. Once audiences know that synthetic media exist, the possibility of fabrication can be invoked to discredit authentic evidence, allowing political actors to escape accountability not by making a falsehood convincing, but by making damaging evidence harder to establish authoritatively as genuine. Synthetic media can threaten not only the reliability of individual pieces of political evidence, but the shared capacity of a political community to establish what counts as authentic evidence at all.

This dual dynamic was visible in Serbia in November 2024. In the aftermath of the Novi Sad railway station canopy collapse, an audio recording surfaced that appeared to capture a ruling-party official instructing activists to provoke incidents during commemorative protests. It was swiftly dismissed as possibly AI-generated. On the same day, a pro-government portal published an openly synthetic recording of an opposition politician to demonstrate how easily convincing audio could be fabricated. Synthetic media were therefore used not only to create political content, but as a resource for casting doubt on separate, politically damaging evidence. The public dispute shifted from what the original recording revealed to whether it could be trusted at all.

This defensive use of AI emerged in an environment where synthetic political content was already visible. Since August 2023, AI-manipulated videos targeting opposition politicians had appeared in Serbian political communication, including on national television. Deepfake-related manipulation can operate in two directions: offensive fabrication, which manufactures false evidence, and defensive dismissal, which invokes the possibility of AI manipulation to contest damaging evidence.

The defensive dimension of deepfake awareness has been conceptualised through the “liar’s dividend”, whereby political actors can exploit the perceived possibility of synthetic manipulation to dismiss damaging authentic evidence as fabricated (Chesney and Citron 2019). However, existing scholarship has paid less attention to how fabrication and dismissal may interact, particularly in political environments characterised by media capture and

uneven institutional safeguards. This thesis addresses that gap by asking: How do deepfake-related truth-subversion practices challenge epistemological security in Serbia?

This question is deliberately relational. Whether these practices actually challenge epistemological security is not settled by the practices alone; it depends on how institutions and media respond, whether contested evidence is disseminated, verified and authoritatively resolved. The same fabricated or disputed recording may destabilise shared standards of truth in one institutional setting and be quickly contained in another. Analysing the practices requires analysing the institutional conditions that determine their effect.

Serbia provides a relevant setting because both forms of deepfake-related practice appeared within the same political and media system during a compressed period of electoral competition and sustained anti-government mobilisation. Although Serbia's exact regime classification is contested, it retains formally democratic institutions while experiencing substantial democratic backsliding, media capture and weakened institutional checks. The thesis therefore uses "hybrid regime" as a broad descriptor of Serbia's political and institutional setting, while "democratic backsliding" and "autocratisation" refer to the process through which that setting has evolved. The analysis does not depend on Serbia fitting a single regime typology.

The study adopts epistemological security, following Adler and Drieschova (2021), as its analytical framework. Rather than asking whether individual citizens believed particular deepfakes, it examines whether shared and credible procedures remained available for determining whether contested political evidence was authentic or fabricated. The analysis is located at the collective level and, consistent with Adler and Drieschova's supply-side orientation, focuses on the political actors, media structures and institutional conditions through which truth-subversion practices are produced, disseminated and contested. Importantly, disagreement or an unresolved authenticity dispute does not by itself demonstrate epistemological insecurity. The relevant question is whether recognised procedures are capable of verifying contested evidence and producing an authoritative resolution.

Methodologically, the thesis uses a qualitative single-case study and theory-guided qualitative document analysis. It reconstructs and compares seven deepfake-related episodes in Serbian politics between August 2023 and August 2025 using media reporting, institutional and legal documents, expert and fact-checking assessments and documented

public statements. These include the 2023 AI-manipulated videos targeting opposition figures broadcast on national television; a 2024 fabricated statement attributed to an election candidate; a 2024 AI-generated video of the then-prime minister; a 2024 recording of a senior ruling-party official that was dismissed as AI-generated; a synthetic audio imitating an opposition politician published the same day; a 2025 contested recording involving two opposition figures; and a 2025 AI-generated video targeting a lawyer associated with the protest movement. The episodes are treated as diagnostic units through which differences in fabrication, dismissal, verification and institutional response can be compared. The study does not seek to measure audience beliefs or society-wide levels of trust.

The thesis makes three main contributions. First, it shows that offensive fabrication and defensive dismissal can operate as connected practices, with a directly documented strategic connection in one of the episodes. Second, it demonstrates that the epistemological significance of deepfakes depend not only on technological sophistication, but on the relationship between credible verification, institutional uptake and dissemination power, as credible verification may exist without becoming publicly authoritative. Third, the thesis develops the concept of selective epistemological security, referring to the politically uneven activation of institutional procedures through which contested evidence is authenticated and truth claims receive authoritative resolution. Across the examined episodes, institutional responses varied markedly depending on the political context, although the study does not claim to establish institutional motive or causality.

The thesis proceeds in five chapters. Chapter 1 reviews the literature on AI, deepfakes, hybrid regimes and media capture. Chapter 2 develops the theoretical framework and operationalises epistemological security into four analytical dimensions. Chapter 3 presents the research design and methodology. Chapter 4 analyses the seven deepfake-related episodes, while Chapter 5 discusses their implications for truth subversion, epistemic authority and selective epistemological security. The conclusion summarises the findings and identifies directions for future research.

1. Literature Review

Artificial intelligence (AI) has increasingly expanded beyond its initial association with military and governance applications into the domain of political communication, introducing new forms of influence and manipulation. While earlier digital technologies enabled the rapid dissemination of misleading or selective information, recent developments in AI have transformed not only the scale but also the nature of such practices. Among these, deepfake technology represents a particularly significant shift. Unlike previous forms of data-driven manipulation, which primarily relied on targeting or amplifying existing narratives, deepfakes enable the creation of highly realistic synthetic audio-visual content that can depict events that never occurred. As a result, they not only contribute to the spread of false information but also challenge the distinction between authentic and fabricated content itself.

This chapter reviews the existing scholarship on the role of AI in IR and political communication, with a particular focus on the emergence of deepfakes as a tool of political manipulation. It first examines how scholarship in IR has evolved from a focus on military and strategic applications of AI toward a broader consideration of its role in information environments and political communication. It then analyses the relationship between AI, political communication and democratic trust, before turning to the specific challenges posed by deepfakes as a form of synthetic media. The discussion subsequently situates these developments within the context of hybrid regimes and digital authoritarian practices, showing how such environments shape the use and impact of AI-enabled manipulation. The chapter concludes by identifying gaps in the literature and outlining the relevance of the present study.

1.1 AI in IR: From Military Applications to Information Environments

AI is increasingly understood not simply as another digital innovation, but as a structural transformation in International Relations (IR) that reshapes how power, governance and political interaction are organised. Unlike earlier technological developments, AI affects not only the tools of political action but also the underlying conditions under which decisions are made and information is processed. The study of AI in IR has gained prominence over the

past two decades, particularly in the 2020s, as generative AI tools and broader governance debates have brought the issue to the forefront of global politics.

Early engagement with AI within IR was primarily shaped by concerns about warfare, military innovation and shifts in the distribution of power. Much of the initial scholarship focused on autonomous weapon systems, lethal robotics and the implications of AI for deterrence and strategic stability. In this context, Horowitz argued that AI could alter the balance of power by providing military advantages to states capable of developing and integrating these technologies, while also accelerating decision-making and targeting processes in ways that could affect deterrence and strategic stability (Horowitz 2018, 54).

Over time, however, the field has broadened beyond military competition and battlefield applications. Reviewing the field, Bode (2024, 68-70) identifies four major thematic strands around which contemporary research has organised: the balance of power, disinformation, governance and ethics. The balance-of-power literature examines how AI capabilities may generate strategic advantages and asymmetries between states, while governance-focused scholarship explores international regulatory efforts and debates surrounding the development and use of AI technologies. A third strand addresses ethical questions concerning accountability, human control and the legitimacy of algorithmic decision-making.

For the purposes of this thesis, however, the most relevant area of inquiry concerns disinformation and the broader effects of AI technologies on political information environments. As both Bode (2024) and Baele et al. (2024) observe, increasing attention is being devoted to how AI technologies reshape the production, circulation and contestation of political information. This shift extends AI scholarship beyond military applications and toward the informational dimensions of contemporary security and political life.

Despite this broadening, IR scholarship on AI remains unevenly developed. Military applications, strategic competition, governance and ethics have received considerably more systematic attention, while the disinformation and information environment strand remains less fully theorised in IR. This gap is significant because the manipulation of information environments has become a central feature of contemporary political conflict, with implications that extend beyond traditional security frameworks. The following section turns to scholarship on AI and political communication in order to examine how these dynamics operate within democratic societies. It focuses on the relationship between AI-mediated

information environments, democratic trust and the conditions under which citizens access, interpret and evaluate political information.

1.2 AI, Political Communication and Democratic Trust

AI can affect democratic self-government by altering the informational conditions under which citizens form political judgements and participate in public life. Access to political information, opportunities to communicate political preferences and exposure to manipulative content are all increasingly shaped by AI-mediated communication systems (Jungherr 2023, 3-4). These effects, however, should not be treated as direct or uniform. Political consequences of changing information environments are mediated by existing social practices, informational shortcuts and other structural conditions through which citizens interpret and use political information (Jungherr 2023, 3-4). Rather than assuming that AI inevitably destroys democratic communication, recent scholarship suggests that its significance lies in the growing opacity of digital information environments and in the need to examine more carefully how these systems influence the quality, visibility and circulation of political information.

Contemporary concerns about democracy are often framed in terms of visible threats such as electoral interference, bots, or disinformation campaigns. However, these phenomena also emerge within broader structural conditions shaped by digital communication environments. The weakening of democratic dialogue cannot be explained solely by intentional manipulation but also by the logic of online spaces themselves, including information overload, the rapid circulation of unverified claims and platform architectures that shape visibility and engagement through algorithmic selection rather than journalistic relevance. In this sense, the problem is not simply that falsehood spreads more easily, but that political communication increasingly takes place in environments that can intensify confusion, polarisation and weakened accountability. Fischer and Jarren (2024, 207-208) describe this transformation as the platformisation of the public sphere, in which public debate is increasingly conducted through private digital infrastructures that do not merely facilitate communication but actively structure its form, reach and effects through their own commercial logics and algorithmic rules.

The democratic risks of contemporary digital communication extend beyond the circulation of isolated false claims, as manipulative campaigns increasingly seek to reshape the broader

conditions under which citizens form opinions, evaluate political actors and participate in public life. Disinformation may distort perceptions of political leadership, suppress or discourage electoral participation, weaken trust in democratic institutions and erode confidence in the news environment itself. Particularly important in this context is the notion of “information chaos,” in which the line between true and false becomes progressively harder to sustain, as well as the related dynamic of “information gaslighting,” whereby the rapid circulation of contradictory or misleading content destabilises citizens’ capacity to assess reality independently (Duberry 2022, 167-171). These dynamics are especially relevant for the study of AI and political communication, since AI systems can amplify the speed, scale, personalisation and persuasive force of manipulative content within already fragile digital communication environments.

While AI amplifies existing dynamics of disinformation, micro-targeting and algorithmic influence, not all forms of AI-enabled manipulation operate in the same way or produce equivalent political effects. Much of the existing literature focuses on the scale, speed and personalization of misleading content, emphasising how AI enhances the efficiency of influence operations.

However, a distinction must be made between forms of manipulation that alter the interpretation of information and those that directly challenge the evidentiary status of information itself. In this respect, synthetic media technologies introduce a qualitatively different type of disruption, as they enable the production of highly realistic audio-visual content that can simulate events that never occurred. Given the historically privileged status of audio-visual material as a relatively reliable form of evidence in political communication, this development raises specific concerns that extend beyond earlier forms of digital manipulation. Rather than simply contributing to the spread of false or misleading claims, synthetic media complicates the ability of audiences to distinguish between authentic and fabricated content, thereby reshaping the conditions under which political information is evaluated.

These concerns are also reflected in attitudes toward AI-mediated journalism. Reuters Institute research finds substantially greater public comfort with human-produced than predominantly AI-produced news: only 12% of respondents reported being comfortable with news produced mainly by AI, compared with 62% for human-made content, while AI-

produced news was also expected to be less trustworthy and transparent (Simon, Nielsen and Fletcher 2025).

1.3 Deepfakes as a Political and Democratic Challenge

Since gaining public attention around 2017, deepfakes have become increasingly difficult to detect because they combine authentic visual material with artificially generated elements. It is a technology which uses real content and digitally manipulates it to depict people saying or doing things that never occurred. It relies on neural networks trained on large datasets to replicate facial expressions, voice and behavioural patterns, enabling convincing face-swapping and speech synthesis (Westerlund 2019, 40). Their impact is amplified by social media environments, where misleading content spreads rapidly and is often perceived as credible, particularly in contexts shaped by information overload and confirmation bias.

At the same time, its growing availability has lowered the barriers to producing manipulated content, as open-source software increasingly enables editing that requires little technical skill (Westerlund 2019, 40). Convincing forgeries, however, still typically depend on large volumes of source imagery, which is why the most realistic deepfakes have tended to target public figures for whom such material is abundant (Westerlund 2019, 41). Deepfakes are treated here not simply as technological artefacts but as political-communication practices whose epistemological significance depends on how they are produced, framed, disseminated, contested and institutionally evaluated.

For conceptual clarity, this thesis uses synthetic media as the broadest category, encompassing audio-visual material generated or modified through AI techniques. Deepfake is used more narrowly for synthetic or manipulated media that convincingly imitate the appearance or voice of an identifiable person. Because the empirical analysis also includes episodes in which political actors claim that potentially authentic evidence is AI-generated, the thesis uses deepfake-related practices as an umbrella term covering both the production of synthetic or AI-manipulated political evidence and the strategic invocation of such technologies in disputes over authenticity. The term therefore does not imply that every episode examined contains a technically verified deepfake.

The intersection of deepfakes and elections became increasingly visible during the 2023 election cycle, when AI-generated content appeared in campaigns across several countries,

prompting growing concern among scholars and policymakers about the vulnerability of electoral processes to synthetic media manipulation (Łabuz and Nehring 2024). A widely discussed example is the 2023 Slovak parliamentary election, where an AI-generated audio recording falsely depicting a leading candidate discussing electoral fraud circulated shortly before voting (De Nadal and Jančárik 2024, 1). The timing of its release, combined with its rapid dissemination during the electoral silence period, raised concerns about its potential influence on voter behaviour.

Comparative cases show that the political significance of deepfakes varies according to timing, purpose and the surrounding information environment. During Türkiye's 2023 presidential campaign, synthetic sexual content targeting presidential candidate circulated shortly before voting and preceded his withdrawal from the race, illustrating the reputational consequences of personalised synthetic-media attacks (Łabuz and Nehring 2024, 460, 466). In Argentina, both major campaigns made extensive use of AI-generated political content, while supporters also began labelling genuine recordings as fabricated, suggesting that widespread synthetic media can coexist with growing uncertainty about authentic evidence (Łabuz and Nehring 2024, 461). In India, a politically damaging audio recording was rejected as potentially AI-generated, while subsequent technical analysis remained inconclusive, illustrating the difficulties that can arise when authenticity itself becomes contested (Łabuz and Nehring 2024, 463).

Electoral timing can intensify the risks associated with synthetic media. Research on generative AI and elections suggests that fabricated content introduced during the closing stages of a campaign may generate politically consequential allegations before journalists, civil-society organisations and technology platforms have sufficient time to assess and challenge them. Such late-campaign interventions, sometimes described as AI-enabled "October surprises," may be particularly difficult to counter when they are disseminated by prominent political actors with substantial reach (Bueno de Mesquita et al. 2023, 4).

Other cases demonstrate the importance of verification and of uncertainty surrounding synthetic media. In Taiwan, fabricated audio targeting a presidential candidate was examined by the national Investigation Bureau, which confirmed that the recording was fake, providing an example of institutional verification resolving an authenticity dispute (Łabuz and Nehring 2024, 464). In Gabon, rumours that a presidential video message was itself a deepfake contributed to wider political uncertainty, showing that the perceived possibility of

fabrication may become consequential even without an established deepfake (Łabuz and Nehring 2024, 466). During the Israel-Gaza war, genuine imagery was also dismissed online as synthetically generated, demonstrating how deepfake awareness can be mobilised to question authentic evidence (Łabuz and Nehring 2024, 467).

Public reactions illustrated the difficulty of assessing the political impact of deepfakes, as debates often oscillate between alarmist claims and arguments that misinformation has only a limited reach. While some studies suggest that exposure to misleading content remains relatively low, this is harder to determine in practice, particularly as disinformation increasingly circulates through encrypted platforms such as Telegram and WhatsApp, where communication is less visible and more difficult to regulate (De Nadal and Jančárik 2024, 4).

At the same time, the influence of deepfakes cannot be reduced to levels of exposure alone. Even content consumed by smaller groups may have disproportionate political effects, especially in environments characterised by low trust in media and public institutions, where individuals are more likely to rely on alternative sources of information. Moreover, political actors themselves may play a central role in amplifying or legitimising such content, further shaping its impact within the broader information environment. These dynamics suggest that the effects of deepfakes are closely tied to the structural conditions in which they emerge, rather than to the technology alone.

Recent scholarship has nevertheless questioned whether deepfakes are uniquely capable of altering political attitudes or behaviour compared to earlier forms of misinformation. In one of the most extensive experimental studies on political deepfakes, Barari, Lucas and Munger (2025) find no distinctive advantage for political deepfakes over other false claims presented through text or audio. Their findings challenge claims that deepfakes possess a uniquely manipulative advantage over other forms of misinformation, rather than suggesting that deepfakes are ineffective or incapable of deception.

The study emphasises that political reactions to deepfakes are strongly shaped by broader factors such as partisanship, digital literacy and pre-existing attitudes. Particularly important is the observation that respondents sometimes perceived authentic videos as fake when the content conflicted with partisan expectations, suggesting that deepfakes may contribute less to mass persuasion than to broader uncertainty surrounding the authenticity of political information (2025, 13-15).

However, the transferability of these findings remains limited. The study was conducted within a controlled experimental setting in the context of the United States as a consolidated democracy with comparatively stronger institutional trust and media pluralism. In contrast, political communication in hybrid regimes and broader autocratising contexts such as Serbia operates within more polarised and fragmented information environments characterised by media capture, weakened institutional credibility and lower trust in traditional sources of information. Under such conditions, the significance of deepfakes may lie less in their direct persuasive capacity and more in their ability to intensify existing informational and democratic instability regarding what can be accepted as authentic political reality.

Beyond empirical doubts about the persuasive power of deepfakes, Habgood-Coote (2023) offers a broader philosophical challenge to what he calls the “epistemic apocalypse” narrative. He contends that recordings have never served as self-sufficient sources of knowledge; rather, their trustworthiness has always relied on established norms governing how they are produced, processed and disseminated. Viewed from this standpoint, deepfakes do not fundamentally destroy the epistemic value of audio-visual evidence. Instead, they represent a further quantitative decline in recording practices that have long been susceptible to manipulation, such as misleading editing, contextual distortion and other breaches of epistemic standards (Habgood-Coote 2023, 6-8).

He also challenges claims that synthetic-media manipulation is historically unprecedented. The history of photography demonstrates that manipulated recordings have existed throughout the development of visual media and that societies have responded, in part, by changing professional standards, sanctions and social practices surrounding their production and reception (Habgood-Coote 2023, 13). He argues that deepfake technology should be understood as a techno-social problem whose consequences depend on the social and political context within which recording practices operate, rather than as the automatic effects of an intrinsically dangerous technology (Habgood-Coote 2023, 18-20). Habgood-Coote's argument therefore shifts the analytical question from whether deepfakes are intrinsically epistemically destructive to the institutional and normative conditions through which audiovisual evidence is produced, verified and sanctioned. That condition, largely outside the scope of a debate conducted with reference to consolidated democracies, is the one this thesis examines.

A growing body of scholarship argues that the primary danger of deepfakes does not necessarily lie in their capacity to persuade audiences of specific falsehoods, but rather in their broader effects on trust and informational certainty. Vaccari and Chadwick (2020) found that exposure to deceptive deepfake videos did not significantly increase belief in false information, but it did increase uncertainty among viewers regarding the authenticity of political content in general. Importantly, participants who were exposed to deceptive deepfakes were significantly more likely to doubt that the content was authentic, even when they were not fully deceived by the manipulated material. This increased uncertainty subsequently reduced trust in news encountered on social media (Vaccari and Chadwick 2020, 7-8). In this context, the significance of deepfakes lies less in convincing citizens to believe a particular fabrication and more in fostering a broader climate of informational indeterminacy, scepticism and withdrawal of trust from both fake and genuine political content alike (Westerlund 2019, 43).

Recent empirical research further supports the argument that the political significance of deepfakes may lie less in successful deception than in the broader uncertainty they generate. Analysing Twitter discussions during the Russian invasion of Ukraine, Twomey et al. (2023) found that users often labelled real footage as artificially generated, while the term “deepfake” increasingly functioned as a broader expression of distrust toward media, political actors and official information sources (Twomey et al. 2023, 12-13). The authors distinguish between “healthy scepticism,” which encourages verification and critical evaluation of information and “deepfake-fuelled scepticism,” in which awareness of synthetic media leads individuals to question the authenticity of political information more generally. Taken together, these studies suggest a growing convergence in the literature: the primary political risk of deepfakes lies less in successful deception than in their capacity to generate uncertainty, scepticism and contestation over authenticity.

The literature reviewed so far primarily examines deepfakes as offensive tools: instruments for producing and circulating fabricated political content. However, an emerging strand of scholarship highlights an equally important but comparatively underexplored dynamic: the defensive use of deepfake awareness. Chesney and Citron (2019) conceptualise this phenomenon through the notion of the “liar’s dividend,” referring to the strategic advantage gained by political actors who exploit public awareness of deepfakes in order to dismiss authentic evidence as fabricated or manipulated (Chesney and Citron 2019, 1785-1786). Pawelec (2022, 19-20) documents multiple cases in which political actors have already

exploited this mechanism, including supporters of US President Donald Trump claiming that politically damaging recordings were deepfakes, demonstrating that the liar's dividend is not only a theoretical possibility but an increasingly observable political practice.

While existing scholarship has devoted considerable attention to the offensive dimension of deepfakes, the defensive dimension may be equally, if not more, consequential for democratic accountability, as it undermines the evidentiary basis on which political claims can be verified and actors held accountable. Importantly, this dynamic does not require the existence of sophisticated deepfakes themselves; the mere public awareness that manipulated media are possible can create opportunities for strategic denial and contestation of authentic recordings. The distinction between offensive and defensive uses of deepfake technology constitutes a central component of the analytical framework developed in this thesis and is elaborated further in Chapter 2.

1.4 Hybrid Regimes, Digital Authoritarianism and Media Capture

While the existing literature on deepfakes primarily focuses on consolidated democracies or broader international security concerns, the political effects of synthetic media cannot be understood independently of the regime environments within which they circulate. The relationship between deepfakes, political communication and public trust is shaped not only by the technological characteristics of manipulated media themselves, but also by the institutional conditions governing the production, circulation and contestation of information.

This section situates deepfakes within the broader scholarship on hybrid regimes and digital authoritarianism to examine how the manipulation of informational environments operates in political systems that combine formally democratic institutions with authoritarian practices. Such contexts are analytically significant because the institutional safeguards expected to constrain information manipulation, including independent media, judicial oversight and regulatory accountability, formally exist but operate in weakened or selectively enforced forms. As a result, manipulated or contested political content may circulate within informational spaces characterised by heightened polarisation, asymmetrical media power and reduced capacity for effective public verification or institutional correction.

Hybrid regimes occupy an intermediate position between liberal democracy and full authoritarianism, combining formally democratic institutions with authoritarian practices.

Rather than representing temporary or incomplete democracies, they are now widely understood as relatively stable regime types whose institutional arrangements persist over time (Levitsky and Way 2010, 2-3).

One of the most influential conceptualisations of hybrid regimes is competitive authoritarianism, which describes systems in which democratic institutions remain formally in place, but the government in power systematically undermines political competition (Levitsky and Way 2010, 4-5). In competitive authoritarian systems, formal democratic institutions such as elections, courts, legislatures and media continue to exist, but they operate under systematically unequal conditions. Officials preserve the appearance of democratic competition while tilting the political field through control of state resources, pressure on opponents, media dominance and the politicisation of institutions (Levitsky and Way 2010, 6-7). Elections are therefore held regularly, but competition is not fair.

This distinction matters for the analysis of deepfake-related manipulation because the institutions that should function as epistemic safeguards, namely independent courts, autonomous regulators and free media, are formally present but often weakened, selective, partially or fully captured. As a result, fabricated content and attempts to dismiss authentic information circulate in an environment where corrective mechanisms already suffer from limited credibility and reduced institutional autonomy.

Digital authoritarianism specifically shows how political power increasingly operates not only by censoring information, but also by shaping the digital conditions under which information is produced, circulated and interpreted. Importantly, digital authoritarian practices are not confined to authoritarian regimes. They may also emerge in democratic contexts, where governments justify them as necessary to combat disinformation, protect public order, or safeguard democracy, even though such practices may ultimately weaken democratic accountability and freedom of expression (Maerz 2026, 164). Furthermore, three forms of digital authoritarian practice are distinguished: manipulative, surveilling and controlling interventions in the online public sphere.

For this thesis, the most relevant category is manipulation, which includes disinformation on social media, the strategic use of “fake news” narratives and the abuse of defamation or copyright laws to influence online discourse and shape official narratives of truth (Maerz 2026, 166-167). In this sense, digital authoritarianism distorts the conditions through which citizens access reliable information, share alternative interpretations and hold political actors

accountable. This is also important for epistemological security because the information environment may already be compromised by manipulative state practices before deepfakes enter it.

In hybrid regimes, manipulation of the information environment often operates through media capture rather than open censorship. As Bieber (2018) shows in the Western Balkan context, contemporary competitive authoritarian regimes rely less on direct state control of the media than they did in the 1990s. Influence is instead exercised through a combination of loyal media outlets with questionable ownership structures, economic pressure on independent media and intimidation or censorship of journalists (Bieber 2018, 347).

1.4.1 Serbia: Democratic Backsliding and Media Capture

Serbia's democratic decline is increasingly analysed through the framework of hybrid and competitive authoritarian regimes, and its media environment sits at the centre of that decline. The structure of media capture documented in this literature: asymmetric dissemination power, weak regulatory constraint and entrenched practices of delegitimation is also the environment into which synthetic media have entered. Three features of that environment are established below: the regime's competitive-authoritarian character, the institutional mechanisms through which media capture operates, and the use of captured media to delegitimise political challengers.

Bursać and Vučićević characterise contemporary Serbia as a hybrid regime with a dominant political party, in which formally democratic electoral competition exists but takes place on a systematically uneven playing field. Incumbents benefit from unequal access to media, public resources and state institutions, and the conditions under which opposition actors can compete are substantially constrained (Bursać and Vučićević 2025, 65). They also explain the regime's continuing advantage through the concepts of nationalisation of elections and stabilitocracy.

Nationalisation allows the SNS to transfer Aleksandar Vučić's personal popularity and media dominance from the national level to provincial and local contests, so that lower-level elections increasingly reflect national political dynamics rather than autonomous local competition. Stabilitocracy captures the external dimension of this system where democratic

violations are tolerated or downplayed by Western partners when the government is seen as cooperative on higher-priority geopolitical issues (Bursać and Vučićević 2025, 74).

Vladislavljević's analysis shows that Serbia's democratic decline was uneven. While competition, participation and accountability were relatively robust between 2001 and 2010, by 2015 competition and accountability had weakened sharply, even as civil-society participation remained comparatively resilient. This produced an increasingly unequal political arena in which opposition actors struggled to challenge incumbents, resembling hybrid-regime politics and raising doubts about Serbia's democratic status (Vladislavljević 2020, 26, 30).

This uneven playing field is produced and sustained through identifiable institutional and financial mechanisms rather than through ownership alone. Krstić (2023) and Dragojlov (2026) identify a similar architecture of indirect media control: political influence over the Regulatory Body for Electronic Media (REM) and its repeated failure to sanction violations by broadcasters; the allocation of state advertising in ways that disproportionately benefit pro-government outlets; and media project co-financing, introduced in 2014, which has developed into a clientelist mechanism for directing public resources towards media close to the governing party.

Although intended to provide transparent financial support for public-interest media content, the co-financing system has been criticised for biased procedures, unclear decision-making and strong state influence over the allocation of funds (Krstić 2023, 9, 11; Dragojlov 2026, 396-400). Together, these mechanisms sustain an unequal media environment in which pro-government outlets command much larger audiences, while critical media operate with more limited reach, strengthening the ability of ruling actors to promote favourable narratives and marginalise opposing voices.

Milutinović and Krstić (2026, 10) show that politically aligned narratives can obtain broad circulation through an interconnected network of supportive broadcast, print and digital outlets, reflecting the wider role of pro-government media in amplifying disinformation. At the same time, much of the corrective work is carried by independent news organisations, fact-checking projects and civil-society actors, which continue to challenge misleading and manipulated content despite limited formal authority and significant institutional and legal constraints (Milutinović and Krstić 2026, 11).

Beyond determining access to information, media also exercise political power through the allocation of voice and the framing of political actors. Research on the 2010 Belgrade Pride shows that actors may be highly visible in coverage while receiving little opportunity to articulate their own positions, and that media framing can privilege controversy and conflict over substantive claims (Krstić, Ejduš and Vladisavljević 2022, 5-9). The political consequence of this structure is not only unequal access to information but also the active delegitimation of actors who challenge political authority. Krasteva and Vladisavljević show that, following the SNS's rise to power, pro-government tabloids and television networks became instruments for portraying opponents as illegitimate or threatening, while broader government narratives increasingly constructed political and civic challengers as enemies or security threats (Krasteva and Vladisavljević 2017, 381, 389-390).

Particularly relevant to this study is the sustained media targeting of opposition politician Dragan Đilas. Vladisavljević and Krstić document a prolonged smear campaign against him by the pro-government tabloid *Informer* during 2013-2014, and place it within a broader pattern in which tabloids and commercial television functioned as political actors in campaigns against opposition politicians, independent journalists, civil-society activists and other government critics. They argue that these indirect strategies of media control can undermine media freedom as effectively as direct control (Vladisavljević and Krstić 2023, 1441-1442).

This mechanism becomes particularly relevant in the context of contemporary mass movements that lack a single political leader who can be individually targeted. Džunić and Golubović show how it operates in coverage of the 2024-2025 student protests. Analysing state-aligned, independent and movement-generated coverage, they describe a captured and asymmetrical media environment in which delegitimation occurs not only through openly hostile rhetoric but also through narrative framing and responsibility attribution (Džunić and Golubović 2026, 2-4). State-aligned outlets portrayed protests as threats to order, socially irresponsible acts or politically manipulated events, with 68% of their analysed coverage classified as delegitimising (Džunić and Golubović 2026, 11).

Taken together, this literature establishes that synthetic media entered an information environment already defined by asymmetric dissemination power, weak regulatory constraint and established practices of mediated delegitimation. Deepfake-related manipulation in Serbia should be understood not in isolation, but against a pre-existing structure in which

political communication, corrective verification and institutional response operate under unequal conditions.

These conditions matter not only because they create unequal opportunities for political communication, but because they may also affect the capacity of state institutions to provide authoritative and broadly credible resolution of contested factual claims. Recent CRTA monitoring has pointed to an increasingly pronounced blurring of the boundary between state institutions and the governing party, alongside growing political polarisation (CRTA 2026).

Where institutional authority becomes closely entangled with governing-party interests, formal competence is not the same as authority. A regulator or prosecutor may have the legal power to examine a disputed recording, yet its ruling may not be accepted as a shared reference point once institutions are seen as aligned with one political side. For epistemological security, institutional autonomy matters not only for democratic accountability, but also for whether contested political facts can acquire credible and broadly authoritative validation.

1.5 Research Gap and Contribution

The literature reviewed in this chapter identifies three relevant but still insufficiently connected bodies of research: AI in IR and political communication, deepfakes as a democratic challenge and digital authoritarianism in hybrid regimes and autocratising contexts. Each addresses one part of the problem, but none fully explains how deepfakes operate within fragile informational environments where institutional safeguards are already weakened. The deepfake literature has largely focused on offensive uses of synthetic media, especially the production and circulation of fabricated content and has often examined their effects through experimental or hypothetical scenarios in consolidated democracies. It has paid less systematic attention to the defensive dimension of deepfake awareness, particularly the liar's dividend and even less to contexts where courts, regulators, media and fact-checking mechanisms operate selectively or under political pressure.

At the same time, the literature on digital authoritarianism shows how governments manipulate digital information environments, but rarely treats synthetic media as a specific political tool. Similarly, AI and IR scholarship has increasingly moved beyond military applications toward information environments, but has not sufficiently examined how AI-

generated content affects domestic standards for evaluating truth and authenticity. The literature reviewed here has not systematically examined how deepfakes operate both offensively and defensively where verification practices and public institutions remain formally present, but credible verification receives uneven institutional uptake and public reach, and authenticity disputes are not consistently resolved authoritatively. Such conditions are directly relevant to epistemological security because they raise the question of whether shared procedures for establishing and validating political facts continue to function as genuinely common procedures across the political field.

This thesis addresses this gap by examining how deepfake-related truth-subversion practices challenge epistemological security in Serbia between 2023 and 2025. By applying Adler and Drieschova's concept of epistemological security, developed in Chapter 2, to a case in which both offensive fabrication and defensive dismissal are visible, the study connects the literature on deepfakes with scholarship on hybrid regimes and digital authoritarianism. Serbia is particularly relevant because both dynamics appear within the same political system and within a relatively compressed timeframe, making it possible to analyse how they interact in practice.

The Serbian case is also theoretically relevant for assessing the portability of the epistemological-security framework itself. Adler and Drieschova develop their account of truth subversion primarily in relation to challenges facing the post-2016 Liberal International Order (LIO) and explicitly leave open whether their analytical framework applies across other geographical and political settings (2021, 15). This thesis takes up that question by applying the framework at the domestic level to Serbia, where institutions of truth production and verification formally persist but operate under conditions of media capture and uneven institutional autonomy.

2. Theoretical Framework

This chapter develops the theoretical framework for examining how deepfake-related truth-subversion practices challenge epistemological security in Serbia. It proceeds in six steps: first, it traces the conceptual move from ontological security to epistemological security; second, it defines epistemological security and adjacent terms; third, it specifies the level of analysis at which the framework operates; fourth, it identifies truth subversion as the mechanism through which deepfakes can weaken the conditions underpinning epistemological security; fifth, it examines the role of epistemic authority and institutional conditions; and finally, it operationalises the framework into four analytical dimensions that structure the empirical analysis in Chapter 4.

2.1 Ontological Security as a Conceptual Starting Point

The concept of ontological security, as developed by Anthony Giddens, refers to “a sense of continuity and order in events, including those not directly within the perceptual environment of the individual” (Giddens 1991, 243). This sense of stability is not given, but constructed through the repetition of everyday routines and practices that allow individuals to manage potential dangers and uncertainties. Through these routinised patterns of behaviour and thought, individuals develop what Giddens describes as a “basic security system,” which helps contain anxiety and maintain coherence in their perception of reality. Importantly, anxiety in this framework is not simply a reaction to external threats, but relates to deeper, often unconscious fears concerning the disruption of this security system and, consequently, the integrity of the self (Giddens 1991, 44-45). In this sense, ontological security can be understood as the condition that enables individuals to experience themselves as continuous beings within a predictable and cognitively ordered environment.

Mitzen extends the concept of ontological security to IR by arguing that states, like individuals, seek to sustain a stable and continuous sense of self over time (Mitzen 2006, 342-343). At the state level, this involves maintaining identity coherence as a precondition for agency. Without a sufficiently stable cognitive environment, actors are unable to organise action or pursue goals consistently, which is why they rely on routinised practices to generate predictability. An important implication is that states may become invested in harmful

patterns of behaviour, since the predictability these routines provide can be valued more than the uncertainty associated with change. In this way, ontological security helps explain the persistence of established relationships, whether cooperative or conflictual, as such patterns tend to resist transformation over time (Mitzen 2006, 343).

Building on the Ontological Security literature in IR, Ejodus refines the concept by shifting analytical attention from its maintenance to the conditions under which it breaks down. Drawing on Giddens, he conceptualises ontological security in world politics as the capacity of collective actors to sustain, at the level of both unconscious and practical consciousness, answers to fundamental questions concerning existence, finitude, relations and autobiography (Ejodus 2018, 883-884). His key contribution lies in the concept of critical situations, understood as disruptions that rupture established routines and bring these otherwise backgrounded questions into discursive consciousness, thereby undermining the actor's capacity to bracket out fundamental questions, generating anxiety and eroding the actor's sense of agency (Ejodus 2020, 16). Importantly, for collective actors, such disruptions are expressed through public discourse, as ontological anxiety becomes visible in the ways political actors articulate and contest fundamental questions.

While the ontological security literature provides an important starting point, it does not fully capture the problem this thesis addresses. Its central concern is the continuity of identity: how actors maintain a coherent sense of who they are in conditions of uncertainty. The research problem examined here is related, but analytically distinct. Deepfake misuse in Serbian politics not only threatens identity narratives; it destabilises the shared knowledge systems through which citizens and institutions evaluate what is authentic, fabricated, credible, or deceptive. The key question is not whether Serbia can maintain a stable sense of collective self, but whether Serbian society can sustain common standards for distinguishing truth from falsehood. Ontological security asks how actors preserve identity continuity; epistemological security asks how actors know what they know together. For this reason, the thesis moves beyond ontological security without dismissing it, using it as a conceptual bridge toward the more precise framework of epistemological security.

2.2 Epistemological Security: Definition and Conceptual Boundaries

Adler conceptualises epistemological security as a condition in which individuals and communities can orient themselves within a shared reality because they regard the knowledge

underpinning it as sufficiently justified and trustworthy (Adler 2019, 290). Post-truth practices threaten this condition by undermining the status of facts and the shared basis on which knowledge claims are validated. For Adler, the consequences extend beyond disagreement over individual facts: epistemological insecurity can obstruct political, social and economic consensus, deepen social divisions and weaken the role of scientific knowledge in public decision-making (Adler 2019, 290-292). It also reaches the practices on which democratic life depends. Because democracy relies not only on representative institutions but on shared deliberation, negotiation and experiential learning, the erosion of a common epistemic basis undermines the very practices through which democratic publics form judgements and act together (Adler 2019, 293-294).

Adler and Drieschova (2021) further develop the concept in relation to truth subversion: the deliberate weakening of the shared standards through which societies determine what counts as credible knowledge (examined in detail in Section 2.4). On their account, epistemological security is sustained by four tightly interrelated factors: practices of truth production, normative understandings of the values and purposes of knowledge production, epistemic authority and dissemination technologies (Adler and Drieschova 2021, 7-8).

Epistemological security is a particularly suitable framework for analysing deepfake misuse, because deepfakes do not only introduce false content into public discourse, they also undermine the very standards by which authenticity, evidence and reality are evaluated. Understanding epistemological security does not require resolving philosophical debates about what truth ultimately is. Rather, it concerns whether a society has a recognised basis for establishing and evaluating truth claims. Such a shared notion of truth is necessary for communication and coordination, because actors need some common ground from which they can assess what is credible, factual and real (Adler and Drieschova 2021, 5).

At the level of social order, the erosion of this common epistemic ground is also connected to political polarisation. Adler and Drieschova argue that truth subversion can deepen divisions to the point that different political camps no longer share a basis for evaluating truth claims, making reasoned discussion across those camps increasingly difficult (2021, 20). Polarisation is relevant to epistemological security not simply because political actors disagree, but because such disagreement can lead to epistemological insecurity when competing groups cease to rely on sufficiently common standards for assessing factual claims.

A similar concern appears in recent work on the epistemological dimensions of populism, where Nawrocki (2024, 445) argues that declining trust in established knowledge authorities can turn political and ideological polarisation into “epistemological ruptures”, as opposing groups increasingly rely on different standards for deciding what counts as factual and true.

Drawing on their account (Adler and Drieschova 2021, 7-11), four main factors epistemological security depends on can be understood as follows. Practices of truth production are the socially recognised methods through which truth claims are produced, evaluated and communicated. Epistemic authority refers to actors socially regarded as legitimate and trustworthy in establishing truth claims. Dissemination technologies shape how truth claims circulate and who can access them. Normative understandings are the shared beliefs about the values and purposes of knowledge production that frame the other three.

These factors are interrelated, so changes in one can affect how the others function. If, for example, a recognised authority verifies that a recording is authentic, that determination stabilises truth only if it can be disseminated widely enough to reach the public and only if audiences still regard such authorities as legitimate; where dissemination is dominated by actors hostile to the finding, or where the authority's legitimacy has already been eroded, even a correct verification fails to settle the dispute.

Building on this account, the thesis treats epistemological security as concerning the stability of the shared practices, standards and authorities through which competing truth claims can be evaluated, rather than the accuracy of every individual belief. Disagreement or error does not in itself constitute epistemological insecurity. The relevant question is whether recognised procedures remain capable of evaluating, correcting and authoritatively resolving contested claims.

The same distinction applies to political deception. Adler and Drieschova distinguish ordinary lying, which continues to presuppose a recognised notion of truth, from truth-subversion practices that attack the standards through which truth claims are established. When a lie is exposed, the wider epistemic order can still recover so long as society retains accepted standards through which the deception can be identified and corrected (Adler and Drieschova 2021, 13). The analytical threshold is therefore not deception itself, but whether the practices through which deception can be identified, challenged and corrected are being weakened.

In this thesis, “shared” does not imply unanimous agreement or universal public trust in a particular institution. It refers to standards, practices and authorities that are intersubjectively and institutionally recognised as legitimate means of evaluating contested factual claims. Formal legal authority can contribute to the public force of a determination, particularly in the case of courts, regulators and prosecutorial institutions, but it should not be equated with epistemic authority. Whether a determination can function as a shared reference point also depends on credible verification, institutional legitimacy and sufficient dissemination.

For the purposes of this thesis, these four factors are not examined equally. Practices of truth production and epistemic authority constitute the primary analytical focus because they can be directly observed through the production, verification and contestation of political evidence. Dissemination technologies are examined through the media and digital channels via which deepfake content circulates. Normative understandings provide the broader democratic context within which truth-subversion challenges accepted standards for evaluating political claims. Together, these four factors provide the conceptual foundation for the analytical dimensions developed in Section 2.6.

Adler and Drieschova’s empirical application operates primarily at the level of the LIO, but they explicitly leave open the applicability of their framework to other geographical and political settings and call for research on the domestic institutional conditions under which truth-subversion practices flourish (2021, 14, 26). This thesis takes up that invitation by applying the framework at the domestic level in Serbia, where institutions and procedures for evaluating political evidence formally persist but operate under conditions of media capture and uneven institutional autonomy. The analysis extends the framework from international order to an already contested domestic epistemic environment without assuming that Serbia began from a fully functioning system of truth evaluation.

2.2.1 Distinguishing Epistemological Security from Adjacent Concepts

Before applying epistemological security to the Serbian case, it is necessary to clarify how the concept differs from adjacent terms. The literature uses “epistemic security” and “epistemological security” in different ways, and the distinction cannot be made on terminology alone. What matters is the object being secured. The section therefore distinguishes Adler and Drieschova’s epistemological security from several related uses of epistemic (in)security.

Seger et al.'s concept of “epistemic security” concerns the robustness of the wider social epistemic infrastructure. They define an epistemically secure society as one that “reliably averts threats to the processes through which reliable information is produced, distributed, acquired and assessed” (Seger et al. 2020, 2). Their framework follows information across this entire process, asking whether weaknesses in production, dissemination, access, attention, source evaluation or trust prevent reliable information from supporting well-informed decision-making (Seger et al. 2020, 10-12, 18-19).

Adler and Drieschova’s concept of epistemological security addresses a different problem. Rather than taking a position on the nature of truth, they ask how social orders establish and evaluate truth claims. Epistemological security depends on actors continuing to rely on sufficiently shared standards and practices for doing so. At the level of social order, insecurity emerges when these common reference points weaken and recognised arbiters no longer provide an accepted basis for resolving competing claims (Adler and Drieschova 2021, 5, 19).

The two frameworks therefore ask different questions: Seger et al. whether the information system reliably produces, circulates and assesses information, and Adler and Drieschova whether actors still share a basis for determining what counts as true when claims conflict. This thesis is concerned primarily with the latter: whether contested political evidence can be evaluated and authenticated through shared and credible procedures.

Vorlíček uses “epistemic insecurity” to examine contestation over the authority to produce legitimate security knowledge. Rather than focusing primarily on uncertainty about whether a proposition is true, her framework asks what happens when established knowledge producers lose their taken-for-granted authority and alternative actors challenge their position (2026, 85-87). Such struggles can alter which actors, forms of expertise and knowledge claims are treated as credible within a security field. This differs from Adler and Drieschova’s framework because the reorganisation of epistemic authority is Vorlíček’s principal object of analysis, whereas for Adler and Drieschova epistemic authority forms only one part of a wider set of practices and standards through which truth claims are evaluated. Contestation of a particular authority therefore does not by itself establish epistemological insecurity in the sense used in this thesis.

Abulof uses “epistemic insecurity” in a substantially different sense, referring to uncertainty about a collective’s future existence rather than the reliability of knowledge claims. In his

account, ontological insecurity concerns questions of collective identity, whereas epistemic insecurity concerns whether the people and their polity will continue into the future (2015, 34). The perceived danger can vary from diminished autonomy or political status to threats to the community's continued physical existence (2015, 41-42). What matters is how members of the community themselves understand and articulate these threats, rather than an external assessment of their objective probability (Abulof 2015, 31). Abulof's concept therefore concerns the future viability of a collective, while Adler and Drieschova focus on the shared standards, practices and authorities through which claims about reality are evaluated.

Kurowska departs more substantially from the other approaches because she places epistemic security within ontological security itself. Her concern is with how authoritative knowledge practices can help secure a collective sense of self. Through what she calls "redemptive hegemony," authoritative knowledge practices can reinforce a particular understanding of reality while generating a sense of collective empowerment (Kurowska 2023, 1–3). The object being secured is therefore ultimately collective identity rather than the reliability of an information system or a shared procedure for adjudicating factual claims. Kurowska also draws attention to the possibility that such security may rest on forms of epistemic authority that privilege some interpretive frameworks while excluding alternatives.

Taken together, these approaches differ principally in what they treat as being secured. Seger et al. focus on the functioning of the wider information system; Vorlíček on epistemic authority; Abulof on the future survival and political viability of a collective; and Kurowska on the collective self. Adler and Drieschova, by contrast, focus on the shared standards, practices and recognised authorities through which truth claims are established and evaluated.

This makes their framework particularly suitable for the present study. The central concern of this thesis is not whether Serbia's information environment is reliable as a whole, whether particular knowledge authorities retain their status, or whether collective identity is secure. It is whether deepfake-related truth-subversion practices weaken the shared basis through which contested political evidence can be evaluated and resolved. Adler and Drieschova provide the theoretical framework for identifying that problem; the empirical analysis then examines whether credible verification is available, whether relevant institutions recognise or act upon it, and whether resulting assessments receive sufficient dissemination to become shared reference points.

The distinction from ontological security nevertheless requires one further clarification. Adler and Drieschova treat ontological and epistemological security as interconnected but analytically distinct (2021, 19), and this thesis follows that position. The present study does not examine the possible downstream effects of epistemological insecurity on individual or collective identity; it is limited to whether shared procedures for authenticating contested political evidence remain operative at the institutional and public-discursive level.

2.3 Level of Analysis

The level at which security is sought is one of the central debates in Ontological Security Studies (OSS), and epistemological security inherits it directly: like ontological security, it must specify whether the security-seeker is the individual or the collective before it can be applied. Some approaches reserve ontological-security seeking for individuals and treat collective structures principally as environments through which personal routines and identities are sustained (Kinnvall 2006; Abulof 2009). Others extend the concept to collective agents, including states (Mitzen 2006; Steele 2008; Zarakol 2010; Subotić 2016).

Ejdus and Rečević (2021, 31-32) offer another way of bridging these positions: instead of simply attributing individual psychological properties to a collective, they conceptualise collective ontological insecurity as an emergent outcome of interactions occurring at lower levels. Their approach therefore provides a basis for analysing collective insecurity without assuming that a state experiences anxiety in the same psychological sense as an individual.

Adler and Drieschova explicitly analyse epistemological insecurity at both the individual level and the level of social orders. At the individual level, it takes the form of “disorientation about what is true and real” (2021, 19). At the level of social order, it arises when competing practices of truth production undermine a socially recognised basis for establishing truth claims and the authority of recognised arbiters weakens. Their framework spans both levels without specifying that one must take analytical priority. A single empirical study, however, cannot remain equally on both: it must decide which level its research question and evidence actually concern.

This choice also reflects Adler and Drieschova’s distinction between the demand and supply sides of populism. Demand-side approaches ask why populations become susceptible to populist and illiberal messages, whereas their analysis concentrates on the supply side: the actors who deploy truth-subversion practices and the media and technological conditions that

enable those practices to spread (2021, 3, 11-12). Following this orientation, the present study brackets individual belief and susceptibility and instead examines truth-subversion practices together with the institutional conditions through which contested claims are produced, verified, disseminated and challenged.

This thesis analyses epistemological security at the collective level of the Serbian political community and observes it through institutional and public-discursive practices. The relevant security-seeker is therefore the political community rather than individual citizens. The object being secured is its shared epistemic basis: the standards, procedures and recognised authorities through which contested political evidence can be evaluated and authoritatively resolved. The analysis asks how this shared basis functions across the examined cases, not whether individual citizens believe particular claims.

Empirically, this shared epistemic basis is observed through the actors and procedures involved in authenticity disputes. Experts and fact-checkers can provide specialised verification; courts, prosecutors and regulators can investigate contested evidence, recognise or act upon expert assessments, or contribute to authoritative resolution; and broadcasters and other media shape the dissemination of competing claims and corrective assessments. The analysis therefore examines how these functions interact when authenticity is disputed rather than treating individual perceptions or beliefs as evidence of collective epistemological security.

Locating the study at this level is defensible on three grounds. First, theoretically, the standards and procedures through which contested evidence is evaluated are intersubjective rather than reducible to individual belief. Adler and Drieschova also recognise an individual form of epistemological insecurity, but this thesis brackets that register rather than denying its existence. Applying their framework to the Serbian political community therefore extends its application from the supra-individual level of the LIO to a domestic setting and examines whether its central concern with shared truth-evaluation practices remains analytically useful at that level.

Second, the available evidence is suited to a collective institutional and public-discursive analysis. The cases document expert verification, institutional responses, media dissemination and publicly observable disputes over authenticity. They do not provide evidence of individual-level disorientation or changes in personal belief, which would require survey, interview or experimental data beyond the scope of this study.

Third, this level makes variation across cases analytically observable. The study can compare whether verification procedures and relevant institutional capacities were activated consistently across politically different episodes, without assuming in advance that any variation follows a particular political pattern. Recurring differences are examined empirically in Chapter 4 and interpreted in Chapter 5. The theoretical level of analysis is specified here, while the deepfake-related episode as the empirical unit of analysis is defined in Chapter 3.

2.4 Deepfake-Related Truth Subversion: Offensive Fabrication and Defensive Dismissal

Truth subversion can be understood as the mechanism through which the shared standards underpinning epistemological security are weakened. It refers to political practices, particularly by illiberal leaders, that do not merely circulate inaccurate claims but function to damage the shared rules through which facts, evidence and credibility are recognised.

2.4.1 Truth, Interpretation and Defactualization

A useful conceptual precursor to contemporary accounts of truth subversion is Hannah Arendt's analysis of factual truth and political lying. For Arendt, factual truth is distinct from both opinion and interpretation. Facts concern events in a shared world and are established through testimony and other forms of evidence, while political actors may legitimately develop different judgements about their meaning and consequences. Political plurality does not require agreement about interpretation, but it does depend on a factual ground to which competing interpretations remain accountable. The unavoidable role of interpretation does not justify collapsing the distinction between fact, opinion and interpretation or altering the underlying factual matter itself (Arendt 2006, 232-233).

This distinction provides an important boundary for the concept of truth subversion used in this thesis. Disagreement over the significance, causes or consequences of an established event remains part of ordinary political contestation. A different epistemological problem arises when the factual status of the event or its evidence is itself manipulated or made indeterminate. Arendt makes this distinction particularly clear when she identifies deliberate falsehood, rather than opinion, as the relevant opposite of factual truth. Presenting a false factual proposition as merely another political "opinion" can itself blur the boundary between

factual claims and interpretative disagreement (Arendt 2006, 242). In the context of this thesis, whether a political recording is damaging, fair or politically significant may remain open to interpretation; whether the recording is authentic or synthetically fabricated is a different kind of question concerning the factual status of evidence.

Arendt further distinguishes traditional political lying from modern forms of mass manipulation. Whereas conventional lies generally concealed particular facts or intentions, modern manipulation can seek to reshape the broader factual context through practices such as rewriting history and political image-making. Established facts may be denied or neglected in favour of a representation intended to function as a substitute for factual reality, while modern communication technologies can give the manufactured image greater public visibility than the reality it displaces (Arendt 2006, 244-246). Although Arendt was writing in a very different technological context, this political logic is directly relevant to synthetic media. Deepfakes extend image-making into the evidentiary domain: the manufactured representation may take the form of the audio-visual evidence through which an event itself would ordinarily be established.

Analysis of totalitarian propaganda illustrates the extreme endpoint of this relationship between political power and factual reality. Arendt describes a system in which facts cease to operate as independent constraints and political power seeks instead to subordinate reality to an internally coherent ideological account (Arendt 1958, 350-353). At its most radical, this process produces subjects for whom, in Arendt's words, "the distinction between fact and fiction... and the distinction between true and false... no longer exist" (Arendt 1958, 474). This argument is not used here as an analogy between contemporary Serbia and totalitarian rule. Rather, it identifies the limiting case of a broader epistemological mechanism: political manipulation becomes qualitatively different when it begins to weaken factual reality itself as an independent standard against which political claims can be assessed.

Arendt's later work demonstrates that the underlying problem is not confined to totalitarian systems. Systematic substitution of shifting stories and manufactured images for factual truth did not need to result in the successful acceptance of one coherent lie. It may instead erode the orientation provided by the distinction between truth and falsehood, making the factual status of political claims increasingly difficult to stabilise (Arendt 2006, 249-250). Her related notion of defactualization captures a condition in which political judgement becomes progressively detached from available factual information and organised around constructed

representations and politically useful images. Importantly, this does not imply that truth itself disappears: the problem lies in weakening the capacity of factual reality to orient political judgement, rather than in eliminating factual reality altogether (Arendt 1972, 31, 36).

These arguments provide an important conceptual bridge to Adler and Drieschova's account of truth subversion. The central concern is not political lying as such, but practices that undermine the standards through which factual claims can be distinguished and evaluated. Adler and Drieschova's use of Orwellian doublethink similarly directs attention to the destabilisation of truth-evaluation standards through contradiction rather than simply to the replacement of one true account by one coherent falsehood (Adler and Drieschova 2021, 14; Orwell 2008, 27-28). Synthetic media sharpen this problem because they intervene directly in forms of evidence ordinarily used to establish what was said, shown or done. They can introduce fabricated evidentiary material while simultaneously making genuine material easier to challenge as potentially synthetic. Deepfakes matter for epistemological security not only because they can make individual false claims more persuasive, but because they can make the factual status of political evidence itself an object of contestation.

2.4.2 Truth-Subversion Practices

While Arendt clarifies the broader political significance of weakening a society's capacity to distinguish fact from falsehood, Adler and Drieschova provide a more systematic account of the practices through which this occurs. They define truth subversion not as the mere circulation of false information, but as practices that weaken the collectively shared standards through which truth claims are produced, evaluated and recognised. Among the truth-subversion practices identified by Adler and Drieschova, false speak, double speak and flooding are particularly relevant to the present study and receive the greatest analytical attention. Their broader taxonomy also includes inflammatory speak, alternative facts, misplaced faith and conspiracy theories. These are not treated as primary analytical dimensions, but are considered where they overlap empirically with the deepfake-related practices examined in the case material. Adler and Drieschova explicitly present these practices as non-exclusive and open to variation across contexts (2021, 15-16).

The first practice, false speak, is open, deliberate lying that erodes the idea that claims can be checked against facts, as in the Vote Leave campaign's false claim that Turkey was about to join the EU. The second, double speak, asserts something and takes it back in the same

breath, putting a claim into circulation while dodging responsibility for it, as when Trump said he would freeze WHO funding and, challenged, insisted he was “not saying” he would. The third, flooding, works through message volume and contradiction: by multiplying competing claims, it makes reliable factual orientation harder, a practice Adler and Drieschova associate with illiberal regimes that disseminate high volumes of contradictory content online (Adler and Drieschova 2021, 15-16).

Deepfakes enable two distinct truth-subversion practices, which this thesis terms offensive fabrication and defensive dismissal. Offensive fabrication produces synthetic content that misrepresents what a political actor said or did; defensive dismissal operates in reverse, rejecting authentic or potentially authentic material as AI-generated in order to evade its implications. Both can contribute to weakened epistemological-security conditions, but through opposite mechanisms.

Offensive fabrication can be understood as a technologically intensified form of false speak. In Adler and Drieschova's typology, false speak refers to “deliberate and obvious lying with the aim to subvert the concept of facts” (2021, 2). Deepfakes intensify this mechanism because they do not only state a false claim about what an actor said or did, but manufacture audio-visual material that appears to show the event itself. This gives the falsehood a particular evidentiary force: Chesney and Citron argue that deepfakes make long-standing forms of lying more damaging precisely because fabricated recordings appear credible (2019, 1771-1772). Offensive fabrication is thus false speak amplified by technology, transforming political fabrication into seemingly observable evidence.

Defensive dismissal does not correspond to a single practice in Adler and Drieschova's taxonomy, but may instantiate false speak where authentic evidence is knowingly misrepresented as fabricated and may overlap with double speak where an AI-fabrication claim is simultaneously advanced and disavowed. This simultaneous assertion and disavowal enacts the doublethink structure, applied to the status of evidence: it puts the AI explanation into public circulation while freeing the speaker from any burden of defending it, so that the claim shapes doubt about the evidence without ever being tested. A comparable dynamic has been observed outside electoral politics. During the Israel-Gaza war, genuine imagery was dismissed online as synthetically generated, illustrating how awareness of deepfake technology can itself be used to contest authentic evidence even where no fabrication has occurred (Łabuz and Nehring 2024, 467).

This thesis uses defensive dismissal to name the observable practice of rejecting damaging material as fabricated, while the established concept of the “liar’s dividend” refers to the strategic advantage that this practice may produce, such as reduced accountability or reputational protection (Chesney and Citron 2019, 1785). As public awareness of deepfakes grows, political actors may invoke the possibility of synthetic manipulation to cast doubt on evidence more credibly. This advantage may be especially consequential in polarised environments, since Barari, Lucas and Munger (2025) find that participants were particularly likely to misclassify authentic scandal videos involving elites from their own party as fabricated.

Rini’s concept of the epistemic backstop clarifies why the erosion of audio-visual evidence is so serious for epistemological security. The epistemic backstop refers to the way recordings function as a check on testimony: because a speaker’s claims can later be measured against a recording, the possibility of recorded verification gives people a reason to speak sincerely and carefully in the first place (Rini 2020, 2-4). Recordings, in this sense do not only document events, they underwrite the reliability of testimonial practice more broadly. Deepfakes threaten this function by making recordings themselves contestable: once authentic recordings can be dismissed as fabricated, audio-visual evidence loses part of its authority as an independent check on political claims (Rini 2020, 7-8). The consequence for epistemological security is direct. When recordings no longer serve as a stable anchor for shared reality assessment, public debate becomes more dependent on competing narratives and less able to resolve disputes through commonly accepted evidence (Rini 2020, 8-10).

2.5 Epistemic Authority, Verification Networks and Institutional Conditions

If truth subversion weakens epistemological security by attacking shared standards of truth, the next question is who protects those standards. Epistemic communities matter because shared standards are carried by actors recognised as competent to evaluate specialised claims. Haas defines them as “a network of professionals with recognised expertise and competence in a particular domain and an authoritative claim to policy-relevant knowledge within that domain or issue-area” (Haas 1992, 3). He identifies four characteristics of an epistemic community: shared principled beliefs, shared causal understandings, common standards for assessing the validity of knowledge claims and a common policy enterprise (1992, 3). He further clarifies that standards of validity and a shared policy enterprise may also be found in

other knowledge-based groups; what more specifically distinguishes an epistemic community is the additional presence of shared principled and causal beliefs (Haas 1992, 35).

Haas's concept is useful here primarily for identifying the epistemic function of specialised expertise under uncertainty. The thesis does not classify the Serbian audio analysts, fact-checkers and investigative journalists as a single epistemic community in Haas's strict sense. Although the actors examined here converge around practices of verification, the present study does not systematically reconstruct their principled and causal beliefs or establish a common policy enterprise across these actors. They are therefore treated more cautiously as epistemic actors within what this thesis terms a verification network: a set of actors who apply specialised knowledge and professional standards to contested claims and can thereby produce credible assessments and reference points in authenticity disputes.

Adler and Drieschova use epistemic authority to describe the socially recognised standing of actors that are trusted and regarded as legitimate in establishing truth claims. In the LIO, this authority is distributed across a diffuse transnational community that includes universities, research institutes, think tanks, international organisations, selected NGOs and some multinational corporations (2021, 8). Epistemic authority depends on social recognition and trust rather than following automatically from expertise or formal position.

Recognising epistemic authority does not mean assuming that experts or knowledge-producing institutions are politically neutral, infallible or entitled to determine truth without challenge. In discussing Foucauldian and postmodern critiques, Adler and Drieschova accept the insight that modern knowledge practices are historically produced and may reinforce existing power relations or marginalise alternative voices (2021, 25). For this thesis, the relevant distinction is not between supposedly neutral knowledge and politically interested knowledge. It is between factual claims that remain open to evidence, verification, criticism and correction, and practices that weaken or bypass those procedures.

Epistemic authority should, however, be distinguished from formal institutional authority in the present study. Actors within the verification network, such as experts, fact-checkers and investigative journalists, derive their epistemic credibility from specialised methods, professional standards and contestable practices of verification. In this study, courts, prosecutors and regulators are analysed primarily through a different function. They may possess formal institutional capacities through which contested evidence can be investigated, recognised, acted upon or authoritatively resolved. These capacities can intersect: an expert

assessment may be methodologically credible without receiving institutional recognition, while a public institution may possess formal authority without itself producing the specialised knowledge required for technical verification. Epistemological security may depend on the relationship between credible epistemic assessment, institutional uptake and dissemination capacity.

Adler and Drieschova treat truth subversion as more than the circulation of false statements: it attacks the practices and standards through which truth claims are established and assessed (2021, 4). For the purposes of this thesis, this makes both verification actors and public institutions analytically relevant, but in different ways. Fact-checkers, experts and other verification actors can produce credible assessments of contested evidence, while relevant public institutions can investigate, recognise or act upon such assessments and, where appropriate, provide authoritative resolution.

The conditions described here are not confined to hybrid regimes as a fixed category. They may also arise more broadly in contexts of democratic backsliding and autocratisation, where formally plural institutions and independent actors continue to exist while their autonomy and capacity to act become increasingly uneven. Serbia represents one such context. The mechanism examined here therefore depends less on a precise regime classification than on the persistence of formally plural institutions under conditions of politically uneven authority and institutional activation.

In autocratising contexts, the relationship among these capacities may become particularly uneven. Where the information environment is manipulated, access to reliable information is restricted and accountability weakened (Maerz 2026, 165-167). Garbe and Maerz use “authoritarian informationalism” to describe a broader mode of rule in which digital and non-digital techniques make surveillance, manipulation and information control increasingly central to political power (2026, 2). This broader condition, rather than any single form of information control, shapes whether epistemic authority can function effectively.

Authoritarian informationalism operates at a different level from digital authoritarianism. Digital authoritarianism is a practice-based concept identifying specific forms of information control such as manipulative, surveilling and controlling practices (Maerz 2026, 163), whereas authoritarian informationalism describes the wider condition in which such practices escalate to become central to the exercise of political power (Garbe and Maerz 2026, 1).

Under such conditions, a credible expert assessment can be produced and still fail to settle a dispute: it may reach no prosecutor's desk and no national broadcast, as later occurred with the audio analyses in the Zobenica case, so that a methodologically sound finding does not stabilise the contested claim. The analytical distinction developed here separates the credibility of verification, the activation of relevant institutional capacities and the dissemination of resulting assessments. This distinction is consistent with Adler's broader premise that rules and norms acquire constitutive force through their recurrent enactment in practice rather than through formal existence alone (Adler 2019, 40). Applied to truth evaluation, the existence of general standards is insufficient if the practices through which those standards are implemented are not consistently enacted.

The formal existence of a public institution should therefore not be equated with the effective activation of its relevant capacities. Adler and Drieschova similarly note that international organisations may continue to exist formally while becoming paralysed under conditions of epistemological insecurity (2021, 24). Applied to the present study, the question is whether public institutions activate procedures capable of recognising credible expertise and contributing to an authoritative resolution when authenticity is disputed.

Building on this distinction, deepfake-related truth subversion does not have to confront a fully functioning verification system. Its epistemological effects are likely to depend on how credible verification, institutional uptake and dissemination power interact under pre-existing conditions of media and institutional unevenness.

The relationship between technology and epistemic authority extends beyond deepfakes themselves. Adler and Drieschova identify emerging technologies, including AI, as capable of reshaping practices of knowledge formation and the distribution of epistemic authority (2021, 27). This broader concern is also reflected in scholarship treating AI as a transformation in how knowledge is produced, circulated and validated (Bode 2024, 66–70; Baele et al. 2024, 6–9). As Haas (1992, 3, 14) emphasises, recognised expertise becomes particularly important under conditions of uncertainty. The broader problem is not only whether AI-generated information is accurate, but which actors and procedures are recognised as capable of validating it. The empirical analysis therefore examines not only whether credible verification was available, but also whether it received institutional uptake and sufficient dissemination to contribute to an authoritative resolution of contested evidence.

2.6 Operationalisation: Four Analytical Dimensions

Having established the theoretical framework, the next step is to translate it into observable analytical dimensions for the empirical analysis in Chapter 4. These four dimensions are theory-driven, derived deductively from the concepts developed in the previous sections rather than generated from the Serbian case. Within this deductive structure, however, the analysis remained open to inductive refinement, allowing patterns not anticipated by the framework to emerge and, where analytically significant, to inform the interpretation. Building on Adler and Drieschova's theoretical framework, the following analytical dimensions translate the concept of epistemological security into empirical indicators suitable for analysing deepfake-related political practices in Serbia.

The four dimensions are not weighted equally. The first three: offensive fabrication, defensive dismissal, and verification and institutional response, form the analytical core and are applied consistently across all seven episodes, although their relevance varies between cases. The fourth dimension, documented public contestation over authenticity, is used more cautiously. It captures publicly observable disputes over whether political evidence is authentic, including competing authenticity claims, documented expressions of uncertainty and cases in which such disputes remain unresolved. It does not measure citizens' beliefs, levels of trust or society-wide epistemological insecurity, which would require representative survey, audience or experimental data beyond the scope of this thesis.

Dimensions 1 and 2 operationalise the truth-subversion practices developed in Section 2.4, while Dimension 3 examines the relationship between verification and institutional response developed in Section 2.5. Dimension 4 provides additional discursive evidence about whether disputes over authenticity are publicly stabilised or remain contested. It is treated as an observable indicator of the functioning of shared truth-evaluation procedures, rather than as a societal outcome of the preceding dimensions. The four dimensions, their theoretical basis and the evidence examined under each are summarised in Table 1.

The presence of disagreement over authenticity does not by itself constitute epistemological insecurity. Political actors routinely contest evidence and interpretation. In this study, evidence of weakened epistemological-security conditions is identified where several indicators converge: competing claims about the authenticity of political evidence; the absence, delay or variation in the activation of recognised verification procedures; a misalignment between credible verification, institutional uptake and broader dissemination

capacity; and the persistence of an authenticity dispute without a commonly authoritative resolution. No single indicator is treated as sufficient on its own. Rather, their combination is used to assess whether the shared procedures through which contested political evidence is evaluated are functioning effectively or are being weakened.

The outcome of an authenticity dispute can take three forms: the material may be authenticated, it may be established as fabricated, or its status may remain unresolved. Authentication and debunking both indicate successful truth evaluation where the determination is produced through credible and recognised verification procedures. Epistemological security does not depend on disputed material being genuine, but on the availability of procedures capable of establishing its status. An unresolved dispute is not treated as evidence that the material is either authentic or fabricated, nor does non-resolution by itself demonstrate epistemological insecurity. It becomes analytically significant when competing authenticity claims persist and no credible verification or authoritative institutional procedure resolves between them. In such cases, the analysis asks why the dispute remained unresolved and what this reveals about the functioning of the procedures through which contested political evidence is evaluated.

Table 1. The Four Analytical Dimensions

Dimension	Definition	Theoretical basis	What the analysis examines
Dimension 1: Offensive fabrication	AI-generated or AI-manipulated content produced to create a false political narrative	False speak, intensified by deepfakes' imitation of audio-visual evidence	Who was targeted, what the content fabricated, how it was framed and distributed
Dimension 2: Defensive dismissal	Authentic or potentially authentic political evidence dismissed as AI-generated or manipulated	May instantiate false speak and, where assertion and disavowal occur together, double speak	Statements dismissing evidence as AI, whether verification was offered or ignored, how the claim was treated in media and documented public discourse
Dimension 3: Verification and institutional response	How verification actors and relevant public institutions respond to deepfake-related incidents and authenticity disputes	Epistemic credibility, institutional authority and institutional uptake as components of shared truth-evaluation procedures	Whether credible technical or fact-checking verification was available; whether relevant institutions responded and how quickly; whether credible assessments received institutional recognition or uptake; and whether there was an authoritative determination of authenticity
Dimension 4: Documented public contestation over authenticity	Publicly observable contestation concerning whether political evidence is authentic or fabricated, or whether its status remains unresolved	Epistemological security as the stability of shared standards and procedures for evaluating truth claims; examined as discursive evidence rather than a measure of public opinion	Competing authenticity claims; documented expressions of uncertainty; whether authenticity disputes are publicly stabilised, resolved or remain contested

3. Methodology

3.1 Research Design

This chapter outlines the research design developed to answer the thesis's central question: how do deepfake-related truth-subversion practices challenge epistemological security in Serbia? The study employs a qualitative single-case study design. Serbia was selected as an analytically relevant case because offensive fabrication and defensive dismissal occurred within the same political and media environment during a relatively compressed period, while its combination of democratic backsliding, formal institutional pluralism and media capture provides a particularly useful setting for examining the portability of a framework developed primarily in relation to the LIO.

The Serbian case is also analytically valuable because the fabricated content appeared within the news programming of a licensed national broadcaster, carrying the institutional appearance of factual reporting, rather than solely through the anonymous online networks where most studied deepfakes circulate. This institutional embedding is precisely what the epistemological security framework is designed to analyse, and it remains comparatively rare in the existing case literature.

The seven episodes are treated as diagnostic embedded units rather than as cases intended to establish the societal effects of deepfakes. Disputes over synthetic evidence make ordinarily latent epistemological processes observable: they reveal who claims authority to establish authenticity, which verification standards are mobilised, how competing assessments are disseminated and whether institutions provide an authoritative resolution. Their analytical significance does not depend on demonstrating large-scale persuasive effects. Consistent with Yin's understanding of case-study research, the study aims for analytical rather than statistical generalisation (Yin 2018).

Adler and Drieschova distinguish their approach from demand-side explanations of populism, which examine the social and political conditions that make populations receptive to populist appeals. Their analysis instead adopts a supply-side perspective, focusing on the actors who develop truth-subversion practices and on the technological and institutional conditions that

enable their dissemination. In particular, they emphasise how changes in the media environment, including the reduced gatekeeping role of traditional media and the growing importance of social media, create opportunities for political actors to circulate truth-subverting claims and weaken established epistemic authority (Adler and Drieschova 2021, 3-4, 11-12).

A qualitative approach was selected because the research examines political practices, meanings and institutional responses rather than measuring audience attitudes or causal effects. Where temporal relationships between episodes are identified, they are treated as plausibility claims rather than demonstrated causal relationships.

The temporal scope extends from August 2023, when the first documented AI-manipulated political content examined in this study appeared in Serbian national media, to August 2025, when an AI-generated video targeting a lawyer associated with the student protest movement demonstrated the continued use of synthetic political content. This period encompasses both offensive fabrication and the use of AI-fabrication claims to contest potentially authentic political evidence. The spatial scope is limited to the Serbian political and media sphere.

The unit of analysis is the complete episode rather than the isolated video or recording. Each episode includes the production or publication of the content, its circulation and framing, competing claims concerning authenticity and the responses of political actors, media, experts and institutions. The design allows recurring and divergent patterns to be examined within the same political, institutional and media environment. The analysis centres on seven principal episodes, treated as units through which interactions among political actors, media, experts and institutions can be examined within the same broader case.

3.2 Data Collection

Data collection is based on qualitative desk research across four categories of sources. No primary interviews were conducted. Because the research examines how deepfake-related incidents were publicly framed, circulated and contested, the relevant evidence is the public record itself: media coverage, institutional statements, expert analyses and documented public reactions, rather than private accounts of motivation or behind-the-scenes decision-making.

The primary empirical material consists of media reporting on deepfake-related incidents in Serbia between 2023 and 2025. Reports were drawn from independent and investigative

outlets, including BIRN/Balkan Insight, Istinomer, N1, Raskrikavanje, Danas, Vreme and Nova, selected for their established track record of fact-based reporting and coverage of digital rights issues. In addition, content from pro-government media, particularly TV Pink broadcasts and the Republika portal, is included not as neutral reporting, but as primary material for analysis, as these outlets were directly involved in the production and dissemination of the content under examination.

Material was identified through targeted searches for documented cases involving AI-generated or AI-manipulated audio-visual political content. Episodes were selected based on three criteria: documented involvement of AI-generated or AI-manipulated content, or documented claims of AI fabrication directed at contested material; political significance within the Serbian electoral or protest context; and sufficient publicly available evidence to assess the relevant analytical dimensions.

These criteria concern the presence of AI-related content and its political relevance, not the nature of the institutional response. The sample was not selected on the outcome of interest: it includes episodes in which institutions responded rapidly and consistently, as well as episodes in which verification was delayed, absent or unresolved. This variation in institutional response is itself part of what the analysis examines, rather than a feature built into case selection.

A second category consists of institutional and legal documents. These include official statements and decisions issued by Serbian state bodies, such as press releases from the Special Prosecution Office for High-Tech Crime, communications from the Regulatory Body for Electronic Media (REM) and available court-related materials, including proceedings connected to legal challenges against deepfake content. Government responses to specific incidents are also included. Access to judicial documentation is partially limited, as not all proceedings are publicly available. In such cases, institutional responses are reconstructed through their reporting in the media sources listed above.

The third category comprises expert and fact-checking analyses. This includes technical assessments of disputed content provided by AI and audio specialists cited in Serbian media, as well as monitoring reports from organisations such as FakeNews Tracker and BIRN's Digital Rights Violations Annual Reports. These sources serve a dual role: they provide evidence of the authenticity of specific materials and constitute part of the epistemic response to deepfake incidents. As such, they are analysed not only as sources of verification but also

as actors that contribute to the maintenance or weakening of shared standards for truth evaluation.

Finally, the study draws on publicly accessible expressions of contestation surrounding deepfake incidents. This includes statements by political actors and public figures, reactions documented in media coverage and selected examples of publicly visible online discourse.

Across all four categories, individual episodes were reconstructed by comparing media reporting, institutional documentation, fact-checking reports, expert assessments and public statements. Given the highly polarised Serbian media environment, no single source was treated as authoritative; where possible, factual claims were cross-checked across different types of evidence to strengthen the credibility and reliability of the analysis.

3.3 Analytical Approach

The material was analysed through theory-guided qualitative document analysis, informed by Mayring's deductive approach (2000). First, each episode was reconstructed chronologically by examining the content involved, its production or publication, circulation and framing, competing authenticity claims and relevant expert or institutional responses. The reconstructed episodes were then compared through the four analytical dimensions developed in Chapter 2: offensive fabrication, defensive dismissal, verification and institutional response, and documented public contestation over authenticity. These dimensions functioned as analytical lenses rather than as mutually exclusive categories, since individual episodes could provide evidence relevant to more than one dimension.

For Dimensions 3 and 4, the analysis distinguishes three components of evidence. Verification material consists of expert and fact-checking assessments and is examined for the methods used, the credibility of the assessment and the conclusion reached concerning authenticity. Institutional material consists of the formal responses of state and regulatory bodies and is analysed for whether, how quickly and how authoritatively institutions engaged with contested evidence and whether they recognised or acted upon available verification. Public-discursive material consists of publicly documented authenticity claims, political statements and media framing and is analysed for how authenticity was asserted, contested or dismissed and whether a credible or authoritative determination became a shared public reference point.

The episodes were not treated as legally or factually identical. Comparisons were limited to a common set of questions: who was targeted; through which channels the content or authenticity claim circulated; how authenticity was framed or contested; whether technical verification was available; which relevant institutions responded; how rapidly and extensively they acted; and whether their intervention produced an authoritative resolution of the authenticity dispute. Comparisons were conducted first among episodes involving similar practices: offensive fabrication and defensive dismissal, before broader patterns of dissemination, verification and institutional response were examined across the full set of cases.

To ensure consistency across the seven episodes, the comparison was organised around a common set of case-level indicators. Table 2 provides an overview of the episodes according to the principal deepfake-related practice, the status of available verification and the nature of the institutional response. The table serves as a comparative analytical overview, while the detailed reconstruction and analysis of each episode are presented in Chapter 4.

Table 2. Overview of Deepfake-Related Episodes Included in the Analysis

Date	Episode/Target	Main Practice	Verification Status	Institutional Response
August 2023	TV Pink - opposition politicians	Offensive fabrication	AI-manipulated content documented	Delayed and limited response: REM issued general guidance; judicial protection was ultimately granted in Aug. 2024
May 2024	TV Pink - Savo Manojlović	Offensive fabrication	AI-manipulated content documented	Delayed regulatory response: REM initiated proceedings; no measure had been identified
July 2024	PM Miloš Vučević	Offensive fabrication	AI-generated video acknowledged by the PM's Office	Rapid institutional response: criminal complaint filed; prosecutors sought preservation of account data; no subsequent judicial resolution publicly identified
November 2024	SNS official Damir Zobenica	Defensive dismissal	Ruling-party actors claimed it was AI-generated; independent audio experts assessed it as likely authentic; no authoritative institutional verification	Legal response without authoritative verification: prosecution dismissed the complaint on legal grounds but did not publicly determine authenticity
November 2024	Republika portal - Đilas	Demonstrative fabrication supporting defensive dismissal	Openly identified as AI-generated by Republika	No separate institutional response identified: primarily functioned as a demonstration supporting the dismissal of the Zobenica recording
January 2025	Đilas-Dražić	Contested AI-fabrication claim	Dražić rejected it as AI-generated; pro-government media presented it as authentic	No authoritative institutional response identified: authenticity claims remained unresolved
August 2025	Lawyer Ivan Ninić	Offensive fabrication	AI-generated video documented	Complaint filed, outcome unresolved

3.4 Methodological Limitations

Some methodological limitations should be acknowledged. First, the study is based on qualitative analysis of publicly available empirical material and does not include data collected through interviews, surveys, or focus groups. The analysis relies on documentary evidence rather than on the perspectives of the actors involved. Although these diverse sources allow for triangulation across different types of evidence, they do not provide direct insight into the perspectives of political actors, journalists, media regulators or audiences involved in the analysed events.

The analysis of public contestation and truth evaluation is illustrative rather than systematic. The study does not seek to measure how Serbian citizens broadly perceived deepfake-related incidents or how exposure affected trust at the societal level. Claims concerning public contestation are limited to documented public statements, media reporting and observable patterns of public contestation rather than representative survey or experimental data. While these documented patterns can provide evidence of weakened epistemological-security conditions at the institutional and discursive level, the thesis cannot determine the extent to which these dynamics affect broader public perceptions of truth and authenticity.

A further limitation concerns the temporal scope of the study. The period under examination was still unfolding at the time of writing, with several legal proceedings unresolved and institutional responses continuing to evolve. Additional deepfake-related incidents may emerge beyond the selected timeframe. The findings should therefore be understood as an analysis of an ongoing and dynamic phenomenon rather than a definitive account of a completed process.

4. Empirical Analysis: Deepfake-Related Truth Subversion in Serbian Politics

This chapter examines deepfake-related episodes in Serbian politics between August 2023 and August 2025. The analysis is structured around the four dimensions developed in Chapter 2: (1) offensive fabrication, (2) defensive dismissal, (3) verification and institutional response, and (4) documented public contestation over authenticity. The empirical material clusters around two moments of heightened political contestation: the December 2023 parliamentary elections and the student-led anti-corruption protests that began in November 2024 after the Novi Sad railway station canopy collapse. This timing is analytically significant because deepfake-related practices became most visible when political stakes were especially high. The chapter does not treat these episodes as isolated technological incidents, but as political practices operating within a contested information environment. It begins by establishing the political and media conditions that shaped how deepfakes were produced, circulated, contested and institutionally handled.

4.1 Political, Media and Institutional Context

Serbia provides a particularly relevant case because deepfake-related incidents occurred within a political system where institutional checks were already weakened. V-Dem describes Serbia as a case of prolonged democratic backsliding, with Vučić and the SNS turning the country into an electoral autocracy since 2014 and subsequently harassing the opposition, increasing electoral irregularities and undermining media freedom (V-Dem Institute 2025, 26). Freedom House similarly documents the erosion of political rights and civil liberties under SNS rule, including pressure on independent media, opposition parties and civil society (Freedom House 2026).

The December 2023 elections illustrated the pattern: ODIHR found the campaign shaped by the president's decisive involvement, ruling-party advantages, media bias, pressure on public-sector employees and misuse of public resources (ODIHR 2024). The fragility of Serbia's electoral environment is underscored by the Kofi Annan Foundation's 2026-2027 Electoral

Vulnerability Index, which assessed Serbia as facing the highest risk of electoral violence in Europe after Russia, attributing this to repeated controversial elections, government intimidation, media imbalance and escalating street mobilisation (Kofi Annan Foundation 2026).

The major democracy assessments do not fully agree on Serbia's classification. Freedom House continues to categorise Serbia as a transitional or hybrid regime, while V-Dem classifies it as an electoral autocracy (V-Dem Institute 2025) and the Bertelsmann Transformation Index 2026 ranks it as a moderate autocracy (BTI 2026). Recent scholarship also increasingly employs the language of competitive authoritarianism and autocratisation. Given these competing classifications, the thesis does not treat any single regime label as definitive. It uses “hybrid regime” as a broad descriptor of Serbia’s institutional setting, while understanding democratic backsliding and autocratising as the process through which that setting has developed.

Analytically, what matters is the underlying institutional condition: formal pluralism persists while institutional autonomy, authority and accountability have become increasingly uneven. This is precisely the kind of setting Adler and Drieschova identify as needing further study when they ask under which regimes truth-subversion flourishes and upon which institutional structures it relies (2021, 26); a domestic context in which the institutions that would ordinarily authenticate contested evidence remain formally present but function unevenly.

The Serbian media landscape is central to how deepfake-related practices operate. Television remains politically powerful despite low public trust in news, which the Reuters Institute recorded at 27% in 2025 (Milivojević 2025). This distrust is also behavioural: Serbian audiences actively avoid mainstream news content at the highest rate in the region (Giebel 2025). The journalist Bojan Elek describes Serbia's media freedom as being in long-term decline, with ruling-party control especially strong over television, while independent journalists face legal, financial and physical pressure (Giebel 2025). CRTA’s 2025 monitoring further confirms the centrality of national television, recording Vučić’s exceptionally high visibility on TV, including 411 live appearances during 2025 (CRTA 2026).

At the same time, independent outlets such as N1, TV Nova, Danas and BIRN operate under sustained pressure, with international media freedom organisations documenting physical attacks, death threats, smear campaigns, legal pressure, spyware attacks and widespread

impunity (European Federation of Journalists 2026). This bifurcation matters because fabrication and AI-dismissal claims often circulate in one media sphere, while fact-checking and debunking remain largely confined to another, limiting the corrective reach of verification networks.

The first cluster of deepfake incidents emerged a few months before the December 2023 parliamentary elections, called early in a period of acute political tension. President Vučić dissolved parliament and scheduled snap elections for 17 December 2023 following the May mass shootings and the “Serbia against Violence” protests they triggered. The shootings prompted demonstrations for 27 consecutive weeks against the “culture of violence” in Serbian media and politics (PACE 2024, paras. 8-11). Local and provincial elections, including in Belgrade, were held the same day as the parliamentary.

The campaign took place in a highly polarised environment that PACE described as marked by an “unprecedented” level of negative campaigning, fearmongering, attacks on opposition and journalists and serious media problems (PACE 2024, para. 4). CEPA similarly described the pre-election period as one of “information chaos,” combining online disinformation, harassment and legal intimidation, with pro-government tabloids and coordinated social media networks amplifying ruling-party narratives (Bajić and Baker 2023).

TV Pink is central to this media environment not only because it has national coverage, but because it functions as a key channel through which political narratives reach mass audiences. CRTA’s mapping shows that state influence over Serbian media often operates indirectly, through public funding, state advertising, service contracts and tax-related privileges. In 2020, TV Pink received the largest share of state advertising among national television stations, amounting to 37% of the total, and in 2021 it was identified as the second most popular television station in Serbia after RTS 1 (Srećković et al. 2022, 60, 91). This public reach matters because Željko Mitrović circulated AI-manipulated content targeting opposition figures first on social media and then through TV Pink’s national news programme. The fabricated material moved from online circulation into national broadcast television.

The second cluster of deepfake incidents emerged during the political crisis that followed the collapse of the newly renovated railway station canopy in Novi Sad on 1 November 2024, which killed sixteen people and triggered sustained protests demanding accountability (OHCHR 2025). Alongside a decentralised, student-led movement organised through direct-

democratic plenums and deliberately resistant to co-optation by established political leaders (Student Blockades n.d.), the protests drew in civic initiatives led by academics and public figures, including ProGlas, which called on citizens to gather in memory of the victims (ProGlas 2024). Through 2025 the mobilisation escalated into one of the largest in Serbia in decades, sustained after the resignation of then-Prime Minister Vučević in January 2025 (V-Dem Institute 2025, 26). This composition is analytically relevant because, as the fabrication cases below show, AI-generated attacks adapted to the movement's leaderless form, shifting from opposition party leaders toward the students, lawyers, academics and independent media associated with the protests.

The disinformation environment in 2025 forms the background to the deepfake cases examined in this chapter. FakeNews Tracker reported that, over six months of monitoring protest-related manipulation, it recorded almost 6,500 examples of manipulative content, with pro-government tabloids such as Informer, Novosti and Alo among the most prominent disseminators (Jovović and Janjić 2025, 5). The scale and density of this manipulation are consistent with what Adler and Drieschova (2021, 2) describe as flooding: the saturation of the public sphere with numerous messages that makes reliable information increasingly difficult to distinguish from manipulation.

4.2 Offensive Fabrication Practices (Dimension 1)

This section examines fabrication practices, meaning cases where AI-generated or AI-manipulated content was used to create false material attributed to political opponents or civic actors. It corresponds to the first analytical dimension developed in Chapter 2, mapping onto what Adler and Drieschova term false speak. The cases cover the full period of analysis and show a pattern of escalation. In 2023, fabrication practices mainly targeted opposition party leaders during the election period. By 2024 and 2025, the targets expanded to civic actors connected to the student movement. This shift shows how fabrication adapted to the changing structure of political opposition in Serbia.

The first documented case in this study occurred in August 2023. Željko Mitrović, owner of the pro-government broadcaster TV Pink, introduced an AI-based format that became known as Čoveče, ne ljuti se (Don't Get Angry, Man), framed as entertainment and satire. The videos targeted leading opposition figures, including Marinika Tepić, Dragan Đilas and Vuk Jeremić, altering their voices and statements to make them appear to say things they had not

said (Đorđević 2023). Đilas later described the content as an AI-generated forgery broadcast in Pink's news programmes, adding that many viewers believed he was really saying it (B. N. 2023). The clips also circulated on X, where videos featuring Đilas and former president Boris Tadić reportedly reached more than 100,000 views before being broadcast in TV Pink's National news on 12 and 16 August (Radio Slobodna Evropa 2023).

Although Mitrović announced the format on his social media accounts as entertainment and satire, stating that all content was fictional (Đorđević 2023), the clips were subsequently broadcast within Pink's news programming. The news anchor introduced the segment as a new satirical program by Mitrović, but the clips themselves were then played without any indication that they had been generated or altered using AI (Baletić 2024). SNS MP Nebojša Bakarec then posted the Tepić video on YouTube and X, likewise without indicating that it was fictional, captioning it with the claim that Tepić was "being honest for the first time" (Đorđević 2023). The satirical designation thus operated primarily at the level of the format: a general framing that introduced the programme, while the fabricated statements themselves circulated through national news programming and political resharing.

Appearing on RTS on 18 August 2023, president Vučić described the Pink videos as crude falsifications, called such use of AI impermissible and unfair, and stated that he expected the competent authorities to take measures (NUNS 2023). Yet no regulatory enforcement proceedings followed. Đilas instead pursued a private legal remedy, requesting a court injunction to prevent further dissemination of the manipulated content. The Higher Court in Belgrade rejected the request in November 2023, but the Court of Appeal subsequently returned the matter for reconsideration (Baletić 2024). Almost a year after the videos first appeared, the Higher Court in Belgrade prohibited Željko Mitrović and PINK Media Group from further broadcasting or otherwise making the AI-generated Đilas video available (N1 Belgrade 2024).

The case also exposed a wider regulatory gap. Tamara Skrozza from Istinomer argued that Mitrović was operating in an "empty space", because Serbia had no clear framework for AI-generated political content (Đorđević 2023). BIRODI director Zoran Gavrilović argued that the targeted politicians should also have filed complaints with REM (Bošnjaković 2023). REM, however, responded only with a public statement restating broadcasters' legal duty not to mislead audiences and recommending visible labelling of AI-generated content; it did not initiate proceedings or impose a sanction against Pink (FoNet 2023; Novaković 2023).

Although the later court injunction shows that judicial intervention was possible, the delayed and limited response meant that AI-manipulated content targeting opposition figures circulated on national television for months without timely sanction.

The timing and targeting of the TV Pink videos are important. They appeared four months before the December elections, during the sustained “Serbia against Violence” protests, and targeted some of the most visible opposition figures. The use of national television also distinguished this episode from many election-related deepfake incidents elsewhere. For example, the Slovak case involved a fake audio clip released shortly before the 2023 election and circulated online during the electoral silence period rather than through a national broadcaster (De Nadal and Jančárik 2024). Comparative research on 2023 election deepfakes similarly shows that most documented cases circulated through social media or online platforms (Łabuz and Nehring 2024).

This pattern did not end with the August 2023 videos. In May 2024, Mitrović created new deepfake content, this time targeting Savo Manojlović, leader of the Kreni-Promeni movement, ahead of the June Belgrade rerun elections in which Manojlović was a candidate for the mayor of Belgrade. The clip showed AI-manipulated footage from Manojlović's appearance on *Utisak nedelje*, altering his voice and statements and broadcasting it on TV Pink's central news programmes (Đokić 2024). The timing again mattered: Pink aired the clip for a day and a half, until the City Electoral Commission registered Manojlović's candidacy, after which campaign-period rules requiring REM to respond within 48 hours would have applied. Because the clip was broadcast before Manojlović's formal registration, the 48-hour campaign-period procedure did not yet apply.

The episode illustrates why timing matters not only for dissemination but also for epistemic correction. Electoral safeguards are temporally structured: the availability and speed of institutional procedures can vary according to when contested content enters political communication. In this case, the timing of the broadcast had the effect of placing the manipulated material outside the faster response procedure that would have applied once Manojlović was formally registered as a candidate.

The broadcast violated precisely the standard REM had articulated after the August 2023 cases and did so despite the Council president's explicit warning that future violations would carry sanctions (Đokić 2024a). REM initially did not react; Cenzolovka noted that, outside the election campaign, the regulator was required to initiate an inquiry within eight days of

learning of a potential violation. On 28 May, the Council eventually initiated proceedings against Pink, with a Council member acknowledging that the fabricated statement could mislead viewers (Đokić 2024b). By mid-July, however, no decision on a measure against Pink in this case had been taken (Predić 2024).

In August 2025, fabrication practices shifted from opposition party leaders to civic actors linked to the student protest movement. An AI-generated video was published on X, targeting lawyer Ivan Ninić, who had become known for providing legal and advocacy support to students and activists facing pressure and arrests (ANEM 2025). The video, titled “Serbian services have proven that every traitor will be discovered and arrested,” visually depicted Ninić in a fabricated scene kissing the Albanian flag and framed him as a traitor and a CIA agent (KoSSev 2025). The same video also targeted independent broadcasters N1 and Nova S, placing them within the same hostile narrative. BIRN’s digital rights monitoring placed the Ninić case alongside other AI-driven smear campaigns in Serbia, highlighting the growing use of fabricated content to target political and civic actors (Reci 2025).

The Association of Independent Electronic Media (ANEM) condemned the video as an unacceptable call for lynching, arguing that Ninić was being targeted because of his legal and public support for students and activists during the protest period (ANEM 2025). The case also raised fear because of its similarity to earlier smear campaigns in Serbia. KoSSev explicitly connected the video to the case of Oliver Ivanović, a Kosovo Serb politician who had been publicly targeted through hostile media narratives before his assassination in 2018 (2025). Ninić announced that he would file a criminal complaint, but also stated that he did not expect the prosecution to react (Lazić 2025).

The Ninić case suggests a possible shift in fabrication practices. Since the student protests were deliberately leaderless and resistant to political co-optation, fabrication practices moved toward the movement’s support infrastructure: lawyers, professors, independent media and other associated public figures. Across the study period, AI-generated attacks adapted to the changing political form of opposition in Serbia.

The episode also overlaps with what Adler and Drieschova describe as inflammatory speak. The synthetic fabrication did not merely attribute false conduct to Ninić, but embedded that fabrication in hostile language portraying him as a “traitor” and “CIA agent” illustrating how false and inflammatory forms of truth subversion can operate simultaneously.

4.3 Defensive Dismissal Practices (Dimension 2)

This section examines dismissal practices, meaning cases where authentic or potentially authentic evidence is dismissed as AI-generated. It corresponds to the second analytical dimension developed in Chapter 2 and to the practice of defensive dismissal. Instead of producing false content, actors exploit deepfake technology to cast doubt on existing evidence and make verification appear uncertain. In the terms developed in Chapter 2, this defensive dismissal may instantiate both false and double speak. False speak because the actor may assert that authentic evidence is AI-generated: a claim that misrepresents the status of the evidence. Double speak because the actor does not need to commit to that claim outright: by simultaneously invoking the AI explanation and distancing themselves from it, they destabilise the criteria by which authenticity is judged without making a falsifiable assertion. This defensive use of AI became especially visible during the student protest period in Serbia.

On 28 November 2024, SSP vice-president Marinika Tepić published an audio recording attributed to Damir Zobenica, an SNS official and vice-president of the Assembly of Vojvodina. In the recording, the voice identified as Zobenica appears to give instructions to party activists concerning the daily memorial blockades held at 11:52, the time of the Novi Sad railway station canopy collapse. The instructions included provoking incidents at the blockades, insulting opposition politicians and public figures, filming the confrontations and sending the footage for media distribution (N1 Belgrade 2024; 021.rs 2024). The timing made the recording politically explosive: it appeared only weeks after the canopy collapse and just as the student blockades were beginning to gain wider national momentum. For the regime, the recording was damaging because it suggested organised provocation at memorial actions that were publicly framed as civic and student-led.

That same evening, the pro-government portal Republika published an openly AI-generated audio imitating opposition politician Dragan Đilas and presented it as a demonstration of how easily synthetic recordings could be produced. The article explicitly described the Đilas audio as false and asserted that both audios represented the same form of manipulation (M.D. Republika 2024). Earlier that evening, president Vučić stated that Zobenica had sent him a message saying it was AI, while adding that he did not know if it was and did not care (Beta 2024). By repeating the explanation while disowning it, Vučić circulated the claim without attaching his own credibility to it and by dismissing the matter as unimportant he shifted

attention from the recording's content, an instance of the liar's dividend (Chesney and Citron 2019).

The AI claim was quickly challenged by technical experts. Dejan Tomka, sound engineer, stated that the Zobenica recording was authentic, while noting that no assessment could be absolutely certain. His analysis focused on the recording's natural speech features: hesitations, sighs, stammering, shifts in tempo and emotional variation, which available AI tools could not convincingly reproduce in Serbian. He also pointed to the consistency of the voice timbre and the fact that the recording sounded like a voice message sent directly to recipients, rather than like synthetic audio (Živić 2024). Experts consulted by Raskrikavanje reached a similar conclusion, arguing that Zobenica did not have enough publicly available recorded material for software to generate such a convincing voice and that Serbian-language voice cloning was not sufficiently developed for this case (Kosanović 2024).

Tomka also compared the recording with the Republika AI-generated Đilas audio, describing the latter as cold, flat and mechanical, which made the natural-sounding Zobenica recording even less plausible as AI-generated (Kosanović 2024). Law professor Miodrag Jovanović publicly argued that the High-Tech Crime Prosecution should examine the case, since forensic verification would be possible if the recording were genuinely AI-generated (Nikolić 2024). No publicly documented forensic investigation followed.

Zobenica later resigned as vice-president of the Assembly of Vojvodina. In February 2025, the Higher Public Prosecutor's Office in Novi Sad dismissed the criminal complaint against him, concluding that the recording contained possible incitement to insult rather than elements of an offence prosecuted *ex officio* (Bokić 2025). The prosecution assessed what the recording called for but did not publicly address the earlier claim that it had been AI-generated, leaving the authenticity question institutionally unresolved. Danas editor-in-chief later remarked that Zobenica's "artificial intelligence" had brought "natural young intelligence" into the streets (N1 Belgrade 2024).

4.4 Verification, Institutional Response and Public Contestation (Dimensions 3 and 4)

This section examines verification and institutional response together with patterns of public contestation across the cases analysed above, corresponding to the third and fourth analytical dimensions developed in Chapter 2. These dimensions are examined together because

institutional responses directly shape the conditions under which public contestation over authenticity occurs; separating them would create artificial repetition without analytical benefit. Rather than reviewing each episode separately, it draws together the cross-cutting patterns that emerged across the study period: institutional asymmetry, regulatory gaps, civil society verification and the public contestation of authenticity. The section shows how deepfake-related practices were shaped not only by technology, but also by the institutional and media environment in which they circulated.

The asymmetry in institutional response is most clearly demonstrated by the Vučević case. In July 2024, an anonymous Facebook account posted an AI-generated video of then Prime Minister (PM) Miloš Vučević. The video manipulated his image and used a synthetic voice to present fictitious government projects (N1 Belgrade 2024). Unlike the deepfakes targeting opposition figures, this case triggered a rapid institutional response. The PM's Office issued a statement, a criminal complaint was filed against an unknown individual and the Ministry of Internal Affairs contacted Meta to request removal of the content and information about the account administrator. The Special Department for High-Tech Crime also requested that the police collect the necessary information and secure data about the account before its removal (Vreme 2024). However, no subsequent arrest, indictment or judicial outcome has been publicly reported.

Ana Toskić Cvetinović, executive director of Partners for Democratic Change, noted that the distinctive feature of the case was the speed of the prosecution's response, which is rare in comparable cases. Similarly, Lawyer Nina Nicović acknowledged that fabricated claims about government projects may warrant a rapid reaction, but noted that Serbian institutions tend to respond promptly to digital-rights violations when politicians are directly affected (Baletić 2024).

The contrast with the cases examined in Sections 4.2 and 4.3 is sharp. AI-manipulated videos of opposition politicians circulated through TV Pink's national news for months, while judicial and regulatory responses were slow, limited or ineffective. The Zobenica case showed a similar pattern from the opposite direction: the prosecution did not resolve the AI claim and later dismissed the criminal complaint. Taken together, the examined episodes suggest a recurring asymmetry in the activation of institutional responses to contested political evidence.

Serbia's legal framework also leaves a significant regulatory gap. There is currently no dedicated binding legislation specifically regulating AI or deepfake content, including its use in political communication. Serbia adopted AI development strategies for 2020–2025 and 2025–2030, as well as the Ethical Guidelines for Development, Implementation and Use of Robust and Accountable Artificial Intelligence in 2023, but these remain primarily strategic and non-binding frameworks rather than dedicated enforceable AI regulation (Government of Serbia 2025; CMS 2026; AI.gov.rs 2023).

This gap is also recognised in recent Serbian scholarship on deepfake regulation, which notes that Serbia has no specific rules on deepfakes and instead relies on existing instruments such as the Criminal Code, copyright law and personal data protection, which remain insufficiently precise for the speed of technological development (Bezbradica et al. 2026, 59). Existing legal remedies are poorly suited to time-sensitive political fabrication, as shown by the Đilas case, where Mitrović's "artistic expression" defence exploited the absence of clear legal boundaries (Baletić 2024).

The regulatory gap adds uncertainty to the institutional response by leaving Serbia without a dedicated framework for AI-generated political content. However, the cases examined here suggest that the central problem is not simply the absence of institutional capacity. Serbian institutions demonstrated that they could mobilise rapidly in some episodes, while comparable capacities were activated more slowly, more narrowly or not at all in others. The regulatory gap is better understood as a condition that can facilitate discretionary and inconsistent responses within an already politically uneven institutional environment.

Where state institutions did not provide authoritative resolution, civil-society organisations, independent media and experts often supplied technical or fact-checking assessments. BIRN documented deepfake incidents and their institutional handling through its digital rights monitoring, including later cases where AI-generated content intensified uncertainty about what was real online (Reci 2025). Raskrikavanje provided a technical assessment in the Zobenica case, concluding that the recording was likely authentic and that current Serbian-language AI tools could not convincingly reproduce the speech patterns in the audio (Kosanović 2024). Istinomer analysed the ethical and regulatory implications of TV Pink's AI-manipulated opposition videos, warning that the format could mislead audiences under the guise of parody (Đorđević 2023). The Crime and Corruption Reporting Network (KRIK)

also monitored the wider disinformation environment, including the limits of AI-based detection and the continued importance of human verification (Karaica 2025).

Together, these actors functioned as the verification network identified in Chapter 2. However, their reach remained limited by Serbia's bifurcated media environment as their assessments circulated primarily through independent media and fact-checking networks, which lacked the national broadcast reach available to TV Pink and other dominant pro-government outlets.

The problem extended beyond cases where independent verification was available. In January 2025, pro-government outlets published an audio recording allegedly featuring opposition politician Đilas scolding his party member over protest organisation. The targeted party member, Maja Dražić, publicly rejected the material as an AI-generated fabrication and described it as part of an SNS smear campaign (RTV Pančevo 2025; Sremskomitrovački portal 2025). No independent expert analysis resolved the dispute.

The Dražić-Đilas recording is the one episode in which no authenticity determination is possible on the available evidence. No forensic assessment was conducted, no institution adjudicated it, and unlike government-aligned media's typical treatment of material damaging to the opposition, it was not widely amplified across pro-government outlets. Its authenticity cannot be established here, and the thesis does not attempt to do so. The episode represents a limiting case in which the epistemic backstop could not operate: no technical assessment established the recording's status and no institutional procedure authoritatively resolved the dispute. The result was an evidentiary vacuum in which competing authenticity claims remained unresolved. The case therefore illustrates the limits of shared truth-evaluation where neither credible verification nor authoritative adjudication is available.

Five interconnected patterns emerge from the cases examined. First, deepfakes function both offensively and defensively within the same political system. They are used to fabricate false material, but also to dismiss damaging evidence as artificial. Second, institutional responses reveal a recurring asymmetry. As examined at the beginning of this section, the state reacted rapidly when the PM was targeted, while responses to cases involving opposition figures, civic actors and government critics were delayed, limited or absent. This pattern persisted even after REM had publicly articulated the applicable standard following the 2023 cases. Third, the satirical characterisation of the AI format created a legal and regulatory loophole.

Although Mitrović presented the format as satire on social media, the fabricated clips reached large audiences through TV Pink's news programme.

Fourth, deepfake deployment clustered around moments of heightened political contestation and adapted to the changing structure of opposition, from opposition leaders before the elections, to civic movement figures and actors supporting the protest movement. Finally, AI-based authenticity claims became a recurring feature of political disputes, although the evidence supports defensive dismissal most strongly in the Zobenica case. AI was no longer only a tool for producing false content, but a ready-made explanation for rejecting inconvenient evidence. Taken together, these developments suggest that deepfake-related practices in Serbia do not operate as isolated technological incidents, but as part of a broader struggle over the credibility of political evidence and the standards through which authenticity is evaluated.

5. Discussion

This chapter discusses the findings in five parts. Section 5.1 examines the interaction between offensive fabrication and defensive dismissal and its implications for the truth-subversion framework; Section 5.2 examines the relationship between epistemic credibility, institutional uptake and dissemination under conditions of media capture; Section 5.3 develops the concept of selective epistemological security; Section 5.4 considers the broader implications of the findings for factual truth, epistemic authority and AI; and Section 5.5 addresses limitations and counterarguments.

The significance of these episodes does not depend on demonstrating that deepfakes were a major driver of Serbia's broader process of autocratisation or that they substantially changed citizens' political beliefs. Their analytical value is diagnostic. Disputes over synthetic and potentially synthetic evidence make visible institutional processes that ordinarily remain in the background: who is recognised as capable of authenticating political evidence, which standards of verification are applied, whether expert assessments receive institutional uptake, which claims obtain high-reach dissemination and whether competing factual accounts receive authoritative resolution. The seven episodes function as stress tests of Serbia's epistemological infrastructure, revealing how shared procedures for establishing political facts operate when authenticity itself becomes contested.

5.1 Offensive Fabrication and Defensive Dismissal as Connected Practices

The findings presented in Chapter 4 suggest that offensive fabrication and defensive dismissal can operate as connected practices within the same field of truth subversion. They can become connected because the prior visibility of AI-generated fabrication can make subsequent claims of fabrication more plausible, although across the broader chronology this is treated as an enabling relationship rather than a demonstrated causal one. Adler and Drieschova (2021) argue that epistemological security is weakened when shared standards for evaluating truth become unstable. In Serbia, this destabilisation operated in two opposing directions within the same political information environment. On the one hand, AI-generated fabrications introduced false content into the public sphere, as illustrated by the TV Pink

videos targeting opposition figures in 2023 and 2024. On the other hand, the Zobenica case showed how awareness of AI-generated manipulation could be mobilised defensively to contest potentially authentic evidence.

The strength of the evidence differs between the Republika-Zobenica sequence and the broader chronology of cases. Within the Republika episode, the connection between synthetic fabrication and defensive dismissal is directly documented: the outlet published an openly AI-generated Đilas audio on the same day and expressly invoked it to support the claim that the Zobenica recording had been fabricated in the same way. Because the Đilas audio was openly labelled, it should not be treated as a conventional case of offensive fabrication intended to deceive. Its function was instead demonstrative: by showing that a politically plausible recording could be synthetically produced, it supplied an example through which the separate dismissal of the Zobenica recording could be made more credible. The connection was strategic in this episode because synthetic fabrication was not merely present in the same information environment; it was deliberately used as a resource for discrediting separate, politically damaging evidence.

Across the broader chronology the thesis makes only a plausibility claim. Earlier fabrications made synthetic political content publicly visible and supplied an available interpretive frame for subsequent AI-dismissal claims, but the evidence does not establish that those earlier fabrications caused the later dismissals or altered audience beliefs.

The Zobenica case illustrates how defensive dismissal can combine elements of false speak and double speak. To the extent that the recording was authentic, the AI claim operated as false speak by misrepresenting the status of the evidence. Even without a conclusive determination of authenticity, Vučić's decision to repeat the AI explanation while distancing himself from it introduced uncertainty into the public dispute over how the evidence should be judged. This resembles double speak: Vučić circulated the AI explanation while simultaneously disclaiming knowledge of whether it was true.

This response also contrasts with Vučić's earlier public position on AI-generated political content. In August 2023, following TV Pink's broadcast of AI-manipulated videos, he described such material as impermissible, unfair and a falsification and demanded state action, yet no institutional response followed. Fifteen months later, during the Zobenica controversy, he publicly repeated an AI-fabrication claim concerning politically damaging evidence. The significance of this contrast lies not in a changed view of AI technology, but in

the different role AI claims played: first invoked to demand action against fabricated content targeting opponents, later to cast doubt on evidence implicating a member of his own party. The effect, whatever the intention, was to shift the debate: rather than Zobenica being asked to explain the recording's contents, the dispute turned on whether the recording itself could be proved authentic.

Deepfake technology gives established practices of denial a new basis by making it plausible to claim that apparently authentic audio-visual evidence was generated by AI. The liar's dividend represents more than another form of political denial. It is a strategic advantage through which the evidentiary basis of accountability can be weakened by exploiting public awareness of AI-generated content (Chesney and Citron 2019, 1785-1786). This is why epistemological security provides an appropriate analytical framework for the present study. The core problem is not simply whether citizens believe false information, but whether shared standards for evaluating authenticity continue to function. The Zobenica case suggests that these standards can be destabilised even when the underlying evidence remains available.

The contested Dražić-Đilas recording from January 2025 illustrates this problem. Pro-government actors presented the recording as authentic; those it targeted rejected it as an AI fabrication and no expert analysis or institutional determination resolved the dispute. The episode illustrates the weakening of what Rini (2020) terms the epistemic backstop: the recording could not function as an uncontested check on competing claims because its own authenticity became disputed and no shared procedure resolved that dispute.

Viewed through Adler and Drieschova's broader taxonomy, the Serbian findings display several overlapping forms of truth subversion. The TV Pink fabrications most closely resemble false speak; the Zobenica episode displays elements of double speak; the high volume of protest-related manipulative content is consistent with an informational environment in which flooding may operate, although the evidence examined here does not establish its effects on audience confusion; and the Ninić case combines fabricated evidence with inflammatory speak. Wider narratives depicting the protests as externally orchestrated may additionally overlap with conspiracy-theory practices, while the examined material provides no equally clear examples of alternative facts or misplaced faith.

This partial correspondence is theoretically significant. It suggests that Adler and Drieschova's framework is portable to the Serbian domestic context at the level of mechanisms rather than as a rigid taxonomy: some practices appear clearly, others overlap

and others are not visible in the material examined. This is consistent with the authors' own treatment of these practices as non-exclusive and open to variation across contexts. More importantly, the Serbian cases show that the epistemological significance of these practices depend not only on their discursive form, but also on the institutional and media conditions through which claims are disseminated, verified and resolved.

5.2 Credible Verification, Institutional Uptake and Dissemination Power under Media Capture

The findings show that the epistemological effects of deepfake-related practices depended not only on whether credible verification was available, but on the relationship among these three capacities. Adler and Drieschova treat dissemination technologies and epistemic authority as closely interrelated: because dissemination structures shape who can communicate truth claims and who can access them, they also influence the distribution of epistemic authority (2021, 7-8). The Serbian case reveals a misalignment between these capacities. Actors with substantial dissemination power did not necessarily apply reliable standards of verification, while actors producing credible assessments often lacked comparable institutional uptake and public reach.

This contrast is most visible between TV Pink and the independent experts and fact-checkers involved in the Zobenica case. TV Pink possessed institutional position and considerable dissemination power. As a licensed national broadcaster, it had mass reach and presented content through a news format carrying an implicit claim to factual reporting. This did not necessarily amount to epistemic authority in Adler and Drieschova's sense, but TV Pink's status as a licensed national broadcaster gave it substantial dissemination power and allowed political claims to appear within an established format of factual reporting.

The AI-manipulated videos targeting opposition politicians illustrate the significance of this position. Unlike deepfakes circulating through anonymous accounts or informal online networks, these videos appeared within the news programming of a licensed national broadcaster. The contrast with Slovakia is instructive: in the Slovak case the disputed deepfake circulated primarily through online networks shortly before the election (De Nadal and Jančárik 2024), whereas in Serbia the fabricated content was carried by a national broadcaster's news programming. The difference was not only one of reach but of institutional form. The videos were broadcast without clear disclosure that the statements had

been AI-generated and some were then reshared by political actors; in one instance, a ruling-party MP reshared a fabricated video captioned as the targeted politician finally speaking honestly, presenting synthetic content as authentic testimony. Although REM stated that AI-generated material in news programming should be labelled and should not mislead audiences, it initiated no proceedings concerning the broadcasts. The dissemination channel therefore did more than increase visibility: it placed fabricated statements within an institutional format associated with factual communication.

The Zobenica case reveals the opposite configuration. Audio experts, fact-checkers and investigative journalists possessed methodological epistemic credibility. Together, these actors operated as part of the verification network identified in Chapter 2: they applied specialised knowledge and professional standards to assess the authenticity of the recording and thereby provided credible reference points in the dispute. Their relevance lies in the epistemic function they performed: reducing uncertainty through technically grounded assessments of contested evidence.

The existence of credible expertise, however, was insufficient to stabilise the authenticity dispute. The prosecution did not meaningfully engage with the expert assessments and, within the material examined, pro-government media did not incorporate them into their coverage. Corrective assessments circulated predominantly through independent media, fact-checking organisations and civil-society networks, without comparable national broadcast amplification, while no public institution provided an authoritative resolution. The Zobenica case shows that an assessment can be methodologically credible without becoming a shared authoritative reference point.

The capacity of verification to stabilise contested evidence is also temporally conditioned. In electoral contexts, a credible correction or institutional determination may arrive too late to stabilise the contested evidence before voting or another politically consequential decision point. The Manojlović episode illustrates this problem from an institutional perspective: the applicable response procedure itself changed according to the timing of the broadcast relative to his formal candidacy.

This misalignment also connects the findings to the literature on media capture and authoritarian informationalism. Maerz (2026) and Garbe and Maerz (2026) show how political power increasingly operates through the manipulation and control of information flows rather than through direct censorship alone, while Bieber (2018) argues that media

control in the Western Balkans often preserves formal pluralism while structurally limiting the influence of independent voices. These conditions can shape both the relative reach of competing truth claims and the institutional uptake available to credible verification. Adler and Drieschova describe the contest between truth-subversion practices and mechanisms such as fact-checking and regulation as a struggle over the epistemological foundations of political order (2021, 4).

Viewed across Adler and Drieschova's four factors, this misalignment extends beyond epistemic authority and dissemination alone. Practices of truth production remained available through expert analysis, fact-checking and investigative journalism, but credible assessments did not always receive institutional uptake. Dissemination was also uneven: some political claims benefited from high-reach national media, while corrective assessments circulated primarily through independent channels. The episodes further suggest contestation over normative expectations surrounding political knowledge, most visibly where synthetic content was defended as satire despite appearing within news formats ordinarily associated with factual reporting. The Serbian case therefore shows that the four factors do not need to deteriorate simultaneously. Weakened epistemological-security conditions may instead emerge when credible truth-production practices, institutional uptake and dissemination structures cease to operate together as a shared system for resolving contested claims.

5.3 Selective Epistemological Security

The Serbian case points toward a specific institutional configuration of epistemological insecurity in which relevant verification capacities and public institutions remain formally available, but the procedures through which contested political evidence is authenticated and authoritatively resolved are activated unevenly across politically different cases. This thesis terms this pattern selective epistemological security. More precisely, selective epistemological security refers to a condition in which formally available truth-evaluation procedures are activated unevenly in ways that affect whether politically consequential authenticity disputes obtain credible verification, institutional uptake and authoritative resolution. "Selective" therefore refers to the uneven operation of collective truth-evaluation procedures, rather than to particular political actors themselves possessing or lacking epistemological security.

This concept refines rather than replaces Adler and Drieschova's account. They recognise that institutions may remain formally present while epistemological insecurity impairs their functioning (2021, 24). The Serbian cases identify a more specific configuration in which epistemic and institutional capacity is neither absent nor uniformly dysfunctional, but unevenly enacted. Following Adler's practice-based premise that rules and procedures acquire force through their recurrent enactment rather than through formal existence alone (2019, 40), the relevant question is therefore not simply whether institutions capable of evaluating contested evidence exist, but whether their truth-evaluating capacities are activated consistently when authenticity becomes politically contested.

Four conditions delimit the concept. First, relevant truth-evaluation capacity must formally exist: experts, verification procedures and public institutions capable of investigating contested evidence must be available. Second, the activation or uptake of these capacities must vary across politically different authenticity disputes. Third, the relevant asymmetry must concern the authentication, verification or authoritative resolution of factual evidence, rather than merely differences in legal sanction or political treatment. Fourth, this uneven activation must affect whether contested evidence can acquire a credible and sufficiently authoritative status capable of functioning as a shared factual reference point. Selective epistemological security is therefore narrower than institutional bias or selective enforcement in general: it concerns specifically the politically uneven operation of procedures through which the factual status of contested evidence is established.

The empirical basis for this refinement lies in a recurring three-part pattern across the examined episodes: rapid institutional mobilisation in the Vučević case; slower, more limited or delayed institutional activation in cases involving fabricated content targeting opposition figures; and the absence of authoritative authentication where contested evidence implicated a ruling-party official in the Zobenica case. These differences do not by themselves establish political causation. Their theoretical significance lies in the fact that they produced different conditions for whether contested evidence could be verified, institutionally recognised and authoritatively stabilised.

The first element of this asymmetry is illustrated by the Vučević case. When an AI-generated video targeted then PM Miloš Vučević, the PM's Office issued a public statement, a criminal complaint was filed, the Ministry of Internal Affairs contacted Meta and the Special Department for High-Tech Crime requested the collection and preservation of relevant

information. Although the available record does not establish a final legal resolution, the episode demonstrates rapid institutional mobilisation when a senior ruling-party actor was targeted.

By contrast, institutional protection in cases involving opposition-targeted fabrication was slower, more limited and, in some respects, absent. When AI-manipulated videos of opposition politicians aired on national television, the Higher Court initially rejected Đilas's request for an injunction. Although the Court of Appeal later overturned that decision and the Higher Court ultimately prohibited further broadcasting in August 2024, the manipulated content had by then circulated for months without timely sanction (Baletić 2024; N1 Belgrade 2024). The regulator did not close the gap: REM issued only a public statement on broadcasters' duties, recommending that AI-generated content be labelled, without initiating proceedings against Pink (FoNet 2023).

The Zobenica case illustrates the third element of this asymmetry in a different form. The disputed recording, which implicated a ruling-party official, was assessed by independent experts as likely authentic, yet it was publicly dismissed as AI-generated by ruling-party officials. No public institution provided an authoritative resolution of the authenticity dispute.

The asymmetry in institutional response cannot be attributed solely to the absence of a political signal: as noted above, the President's own public demand for action against the 2023 fabrications produced no institutional response (NUNS 2023). Across the examined episodes, institutional responses varied across cases involving actors in different political positions, although the evidence cannot establish that political alignment caused those differences. The cases differ in their legal and factual characteristics and cannot establish institutional motive. What they do document is a recurring asymmetry in the speed, extent and form of institutional activation, with consequences for whether authenticity disputes were investigated, resolved and evidentiary credibility defended.

The significance of this pattern lies precisely in the coexistence of formally plural institutions with systematically unequal conditions of operation. Levitsky and Way (2010, 4-7) characterise competitive authoritarian systems as regimes in which elections, courts, legislatures and media remain formally in place but operate on an uneven political playing field. In Serbia, as set out in Section 1.4.1, this unevenness is visible in an information environment where formal media pluralism persists alongside politically structured advantages for pro-government outlets. At the same time, independent media, civil-society

actors and other alternative sources of knowledge have not disappeared; Serbia's democratic decline has itself been uneven, with civil-society participation remaining comparatively resilient even as competition and accountability weakened (Vladisavljević 2020, 26, 30).

Recent research and monitoring reinforce this institutional interpretation. Milutinović and Krstić (2026, 15) describe a structural paradox in which state actors may simultaneously occupy regulatory roles and participate in the dissemination of disinformation, while CRTA (2026) identifies an increasingly blurred boundary between state institutions and the governing party. These conditions matter for epistemological security because formal competence does not itself ensure that institutions can operate as shared arbiters of contested factual claims. Where institutional autonomy is weakened, procedures for recognising credible verification and resolving authenticity disputes may be activated unevenly across politically different cases.

Selective epistemological security is related to, but analytically distinct from what Levitsky and Way (2010) describe as "legal repression": the selective use of formally legal and regulatory instruments against political opponents and critical media. Selective epistemological security identifies a different asymmetry: what becomes unevenly available is not only legal protection, but institutional access to the procedures through which contested evidence is authenticated and evidentiary claims acquire authoritative public validation.

The Zobenica episode shows why selective epistemological security cannot be reduced to "legal repression." A formal legal procedure took place, but it addressed the recording's content without resolving its authenticity. The legal question was adjudicated while the epistemic status of the evidence remained open. This does not mean that the prosecution was legally required to conduct forensic analysis. It means that the procedure did not provide an authoritative basis for determining whether the recording was authentic. Selective epistemological security captures this gap between legal action and authoritative resolution of authenticity.

This unevenness is further reinforced by political polarisation. Adler and Drieschova argue that epistemological insecurity becomes particularly consequential when different parts of society no longer rely on sufficiently shared standards and authorities for evaluating factual claims, making reasoned discussion across political camps increasingly difficult (2021, 20). In Serbia's polarised and bifurcated media environment, uneven verification and dissemination may contribute to conditions in which political actors and audiences encounter

different factual claims, corrective information and epistemic authorities. Under such conditions, even methodologically credible verification may fail to become a shared reference point across the political field. The present study does not establish that Serbian society has fragmented into incompatible epistemic communities, but it identifies institutional and communicative conditions under which such fragmentation may be facilitated.

The literature on digital authoritarianism and media capture helps explain how institutional unevenness is reproduced. Informational control does not need to depend on direct censorship or the elimination of competing voices; it can operate through manipulation, regulatory weakness and media capture that structure the relative reach and influence of different actors (Bieber 2018, 347; Maerz 2026, 164-167). This helps explain how institutions and actors capable of verifying disputed evidence can remain present while differing substantially in autonomy, reach and institutional influence. Digital authoritarianism and selective epistemological security therefore operate at different analytical levels.

In more consolidated democracies, comparatively autonomous institutions and more consistently applied verification procedures may limit the scope for politically selective activation, although they cannot eliminate epistemological contestation or insecurity entirely. Consolidated democracies may combine comparatively autonomous institutions, more consistently activated verification procedures, plural and trusted media systems, and broader societal capacities for evaluating information. None of these conditions guarantees epistemological security, but together they may make shared truth-evaluation procedures more resilient and reduce the scope for politically selective activation. Denmark, Finland, Norway and Sweden combine strong democratic and institutional foundations with high press freedom, comparatively high trust in professional news and substantial digital and media-literacy capacities (Nordicom 2026).

Adler and Drieschova's distinction between ordinary lying, truth-subversion and totalizing truth helps clarify the regime logic developed here. In closed authoritarian regimes, where dissent can be more fully excluded, political authorities may seek to impose a totalizing truth: a dominant account that prohibits competing versions of reality (2021, 12). Whereas totalizing truth imposes one authoritative account, truth-subversion operates differently in contexts where dissent cannot be fully excluded. In democracies and hybrid regimes, where political communication cannot be completely monopolised, it may weaken shared standards by generating confusion among competing claims rather than by imposing a single coherent

account (2021, 14, 17). Hybrid regimes and other autocratising contexts may occupy an intermediate position: competing institutions, media and epistemic actors continue to exist and can challenge official accounts, but they do so under conditions of unequal autonomy, institutional recognition and dissemination power.

Epistemological security should not be treated as a binary condition. Adler's practice-based account allows epistemological insecurity to coexist with institutional contestation and resilience, where courts, media and other communities of practice continue to challenge post-truth practices and develop corrective responses (Adler 2019, 292–294). The Serbian case illustrates precisely this configuration: competing epistemic actors and verification capacities persist, while institutional activation, uptake and dissemination remain politically uneven. Selective epistemological security captures this asymmetry without implying either the disappearance of epistemic pluralism or the imposition of totalizing truth.

5.4 Broader Implications: Factual Truth, Accountability and AI

The Serbian findings illuminate the distinction between factual truth and political opinion developed by Arendt in Chapter 2. Democratic politics necessarily involve disagreement over the interpretation and significance of facts, but such disagreement presupposes a common factual ground to which competing opinions remain accountable (Arendt 2006, 232–233). The Zobenica episode demonstrates a different kind of dispute. The controversy did not remain at the level of how an acknowledged recording should be interpreted or what political significance should be attached to its contents. It shifted to whether the recording itself could be accepted as authentic evidence. Deepfake-related dismissal shifts political contestation from disagreement over the meaning of evidence to disagreement over the factual status of the evidence itself.

Offensive fabrication reveals the opposite side of the same problem. Arendt's analysis of modern political image-making shows how political manipulation can move beyond concealing or misrepresenting individual facts toward constructing representations that compete with factual reality itself (Arendt 2006, 244–246). The TV Pink cases give this logic a specifically synthetic form. The manipulated videos did not simply assert false claims about opposition politicians; they created audio-visual material that appeared to show those politicians making the statements attributed to them. In this sense, deepfakes intensify political image-making by transforming a manufactured political representation into apparent

evidence. The epistemological problem is not only that the content is false, but that the falsehood assumes the form of the very evidence ordinarily used to establish what actually happened.

Taken together, offensive fabrication and defensive dismissal also clarify the relevance of Arendt's notion of defactualization. The narrower problem revealed by the cases is that factual evidence may remain available while becoming less capable of serving as a stable reference point for political judgement (Arendt 1972, 31, 36). In the Zobenica case, the recording remained accessible and was assessed by independent experts as likely authentic, yet the AI claim and the absence of authoritative institutional resolution allowed its factual status to remain politically contested.

The Dražić-Đilas episode presents an even clearer example: competing actors advanced incompatible claims about the recording's authenticity, while no recognised expert or institutional procedure resolved between them. In both cases, the problem was not the disappearance of evidence, but the weakening of its capacity to settle a dispute over what could be accepted as factual. These episodes do not demonstrate society-wide defactualization; rather, they identify institutional and discursive conditions under which factual evidence can lose some of its orienting and adjudicative force.

The stakes of this evidentiary instability are also democratic. Adler and Drieschova themselves observe that established democracies can weaken following elections in which truth-subversion practices play a major role, a dynamic they date to the 2016 US presidential election (2021, 17). Democratic accountability depends on a shared capacity to establish what political actors said and did: elections, oversight and public deliberation all presuppose that citizens can, in principle, distinguish authentic evidence from fabrication and reach authoritative determinations when authenticity is disputed. Where the procedures sustaining that capacity are activated unevenly across politically different cases, accountability itself becomes uneven.

Epistemological security can be understood as one of the preconditions of democratic accountability: the ability to hold power to account depends on the ability to establish, authoritatively and in common, what actually occurred. Epistemological security in democratic contexts consequently also has a temporal dimension: where political decisions occur at fixed points, such as elections, the capacity to establish factual truth may depend not

only on whether credible verification ultimately occurs, but on whether it occurs in time to inform the political process for which the contested evidence is relevant.

Discussion on truth-telling institutions further clarifies why technical verification alone may be insufficient to stabilise factual truth in public life. Arendt assigns particular importance to relatively independent actors such as journalists, judges, historians and other fact-finders whose role is to establish and preserve factual claims against political pressure (Arendt 2006, 251-252). The Serbian cases add an important qualification to this insight. Experts, fact-checkers and investigative journalists were capable of producing evidence-based assessments of contested material, particularly in the Zobenica case, but those assessments did not necessarily receive authoritative institutional uptake or comparable dissemination. Credible knowledge could exist without becoming politically consequential.

The Serbian case illustrates an interaction between technological and political sources of epistemic uncertainty. AI makes the authenticity of evidence easier to contest, media fragmentation and uneven institutional response shape whether those disputes can subsequently be resolved. The challenge does not arise from deepfake technology alone, but from its interaction with the political and institutional environment in which authenticity is evaluated.

The broader implication is that AI does not only create new forms of political content; it also changes the conditions under which contested knowledge claims must be validated. As synthetic content becomes easier to produce and AI-manipulation claims become more plausible, establishing authenticity depends less on the recording itself and more on whether a credible finding is taken up and reaches the public, as the verification-uptake-dissemination relationship in these cases showed. The Serbian cases show that technical assessment alone is insufficient where it does not receive institutional recognition or reach the wider political field. The challenge posed by AI is not simply one of detecting fabrication, but of maintaining procedures capable of converting credible verification into authoritative and publicly consequential determinations.

Deepfakes raise a problem that technological detection alone cannot resolve: who can provide an authoritative determination when the authenticity of politically consequential evidence is disputed? The findings do not imply that any single institution should possess a monopoly over political truth. Rather, they suggest that credible verification, legitimate institutional adjudication and sufficiently broad dissemination must be connected if factual evidence is to

retain its capacity to function as a shared public reference point. Where these functions remain fragmented or are activated unevenly, even accurate expert assessments may fail to resolve political disputes over authenticity. The broader challenge posed by synthetic media is therefore not merely technological, but institutional: preserving factual truth increasingly depends on whether societies maintain credible procedures through which contested evidence can be verified, authoritatively assessed and publicly recognised.

5.5 Limitations and Counterarguments

The strongest challenge to this thesis comes from Barari, Lucas and Munger (2025), who find that political deepfakes do not possess a distinctive credibility advantage over comparable misinformation presented through text or audio. If deepfakes are not uniquely credible relative to conventional misinformation, treating them as a distinct epistemological threat may appear unnecessary. However, the same study finds that participants sometimes misclassify authentic scandal videos involving political elites from their own party as deepfakes, a result directly relevant to the false-positive dynamic examined in this thesis.

The Serbian case suggests that the epistemological threat of deepfakes does not lie primarily in their persuasive power. Rather, widespread awareness of synthetic media gives an established defensive practice, the denial or discrediting of damaging evidence, a new and increasingly plausible technological basis. Experimental designs focused on whether audiences believe fabricated content do not fully capture this defensive dynamic, in which genuine or potentially genuine evidence can itself be dismissed as artificially generated. Barari, Lucas and Munger's findings challenge alarmist accounts of deepfake persuasion while reinforcing the concern at the centre of this thesis: the destabilisation of the standards through which authentic evidence is evaluated. Their study was also conducted in a consolidated democratic context whose institutional conditions differ from Serbia's, whereas the findings presented here suggest that the epistemological significance of deepfakes is partly conditioned by the institutions through which authenticity disputes are verified and resolved.

A second challenge comes from Habgood-Coote's (2023) critique of the "epistemic apocalypse" narrative. If the reliability of recordings has always depended on norm-governed practices of production and sanction, deepfakes should be understood as a techno-social rather than an intrinsically catastrophic problem. This argument qualifies rather than

contradicts the findings of the thesis. The Serbian case similarly shows that the epistemological significance of synthetic media depend on the social and political organisation of recording, verification and sanctioning practices. It also identifies conditions under which adaptation may be constrained: credible expertise existed, but fragmented dissemination, limited institutional uptake and selective enforcement weakened its capacity to stabilise authenticity disputes. The problem lies not in deepfake technology alone, but in its interaction with institutionally uneven practices for producing, disseminating and validating political evidence.

A further theoretical limitation concerns the transfer of the epistemological-security framework to a context different from that for which it was originally developed. Adler and Drieschova examine truth subversion primarily in relation to the LIO and explicitly leave the framework's applicability to other geographical and political settings open. The Serbian case suggests that its central mechanisms are applicable at the domestic level, but also that adaptation is necessary where the epistemological infrastructure is already institutionally fragile and politically uneven. Under such conditions, the relevant dynamic is not only the erosion of shared standards, but their selective institutional maintenance and activation. This thesis addresses that difference through the concept of selective epistemological security. However, a single Serbian case cannot establish that the same refinement applies across hybrid regimes and autocratising contexts generally; comparative research is needed to determine whether similar patterns of politically uneven verification and authoritative resolution recur elsewhere.

As detailed in Section 3.4, the research design does not measure audience-level belief or trust and captures an evolving period between 2023 and 2025. The findings therefore concern institutional and discursive conditions of epistemological security rather than society-wide changes in citizens' beliefs or perceptions. Future research could combine qualitative case analysis with surveys, experiments or longitudinal research to examine how authenticity disputes are received at the audience level and whether the patterns identified here persist as synthetic-media technologies develop.

Conclusion

This thesis asked how deepfake-related truth-subversion practices challenge epistemological security in Serbia's political context. The analysis shows that their significance lies not only in the production of fabricated political content, but also in the relationship between offensive fabrication and defensive dismissal. Offensive fabrication introduces synthetic or manipulated evidence into political communication, while defensive dismissal exploits awareness of such fabrication to cast doubt on potentially authentic and politically damaging material. The two practices can become connected components of a broader truth-subversion dynamic, with a direct strategic connection most clearly documented in the Republika-Zobenica sequence.

The findings do not demonstrate a measurable decline in public trust across Serbian society, nor do they establish how individual citizens interpreted the incidents examined. Rather, they identify weakened epistemological-security conditions at the institutional and discursive levels. Consistent with Adler and Drieschova's supply-side perspective, the findings concern the actors, practices and institutional and media conditions through which truth subversion is produced, disseminated, verified and contested, rather than the demand-side question of why individual citizens accept or reject particular claims.

Across the seven episodes, some authenticity disputes remained unresolved; where credible technical assessments were available, they did not consistently receive institutional uptake; and relevant institutional procedures were activated unevenly across politically different cases. The political significance of deepfakes depended not only on their technical sophistication or persuasive capacity, but also on the relationship between credible verification, institutional uptake and dissemination power.

These findings support the thesis's central conceptual argument. Selective epistemological security, as developed in Chapter 5, identifies an institutional pattern through which conditions of epistemological insecurity may be reproduced even where verification

capacities and formal institutions remain present. Whether this pattern extends beyond Serbia requires comparative research.

Future research should proceed in two directions. Comparative case studies could examine whether the interaction between fabrication, dismissal and selective institutional response recurs in other hybrid regimes and autocratising contexts. Mixed-methods research combining qualitative analysis with surveys, experiments, audience research and computational analysis of platform data could assess how authenticity disputes are received, how widely competing framings circulate and whether the liar's dividend becomes observable beyond individual episodes.

The difficulty of establishing the authenticity of political evidence also points to a broader question about epistemic authority in AI-mediated information environments. As AI-generated content becomes more common, the importance of credible procedures for validating evidence is likely to increase. The Serbian case illustrates this problem in a context where credible verification, institutional uptake and dissemination power were already unevenly activated.

For institutional and policy responses, this implies that measures centred exclusively on detection technology address only part of the challenge. The resilience of shared standards also depends on whether credible verification is available, whether relevant public institutions recognise and act upon it, and whether resulting assessments reach the wider political field. Where these functions remain disconnected or are activated unevenly, improved detection alone cannot prevent authentic evidence from being strategically dismissed.

Deepfakes do not merely add another instrument to the existing repertoire of political manipulation. They reshape the conditions under which political evidence is evaluated, contested and authenticated. In political systems where the institutional infrastructure through which contested factual claims are investigated and authoritatively resolved is already fragile, the central challenge is no longer simply whether individual recordings are believed, but whether political communities retain credible and consistently activated procedures for determining what evidence can be accepted as authentic. Preserving those shared standards is likely to become as important for democratic resilience as identifying fabricated content itself.

References

- 021.rs. 2024. "AUDIO Tepić: Zobenica davao zadatke članovima SNS da izazivaju incidente na akciji 'Zastani Srbijo'." 021.rs, November 28, 2024. <https://www.021.rs/novisad/vesti/394536/audio-tepic-zobenica-davao-zadatke-clanovima-sns-da-izazivaju-incidente-na-akciji-zastani-srbijo>.
- Abulof, Uriel. 2009. "Small peoples": The existential uncertainty of ethnonational communities." *International Studies Quarterly* 53.1. 227-248.
- Abulof, Uriel. 2015. *The mortality and morality of nations*. Cambridge: Cambridge University Press.
- Adler, Emanuel, and Alena Drieschova. 2021. *The Epistemological Challenge of Truth Subversion to the Liberal International Order*. Accepted manuscript, ORCA, Cardiff University.
- Adler, Emanuel. 2019. *World Ordering: A Social Theory of Cognitive Evolution*. Cambridge: Cambridge University Press,
- AI.gov.rs. "Adopted Ethical Guidelines for Safe and Reliable Use of AI." March 24, 2023.
- Arendt, Hannah. 2006. *Between Past and Future: Eight Exercises in Political Thought*. New York: Penguin Books.
- Arendt, Hannah. 1972. *Crises of the republic: Lying in politics, civil disobedience on violence, thoughts on politics, and revolution*. Houghton Mifflin Harcourt.
- Arendt, Hannah. 1958. *The Origins of Totalitarianism*. 2nd enl. ed. Cleveland: Meridian Books.
- Associated Press. "Protesters in Serbian Cities Block Traffic, Standing in Silence for the Victims of Roof Collapse." November 22, 2024. <https://apnews.com/article/serbia-roof-collapse-silent-protest-arrests-83ae578110e0979c30d91368fd7098c7>.
- Association of Independent Electronic Media (ANEM). 2025. "ANEM Alarm: Campaign against Lawyer Ivan Ninić and Targeting Television Stations Nova S and N1 – An Unacceptable Call for Lynching." August 25, 2025. <https://anem.org.rs/en/strane/anem-alarm-kampanja-protiv-advokata-ivana-ninica-i-targetiranje-televizija-nova-s-i-n1-nedopustiv-poziv-na-linc>.
- B. N. "Opposition Leader to Sue Pink TV for Cloning His Voice in Video Clip." N1, August 16, 2023. <https://n1info.rs/english/news/opposition-leader-to-sue-pink-tv-for-cloning-his-voice-in-video-clip/>.

Baele, Stephane J., Iqraa Bukhari, Christopher Whyte, Scott Cuomo, Benjamin Jensen, Kenneth Payne, and Eugenio V. Garcia. 2024. "AI IR: Charting International Relations in the Age of Artificial Intelligence." *International Studies Review* 26.2.

Bajić, Mila, and Grant Baker. 2023. "'Information Chaos' Plagues Serbia Elections." Center for European Policy Analysis (CEPA). <https://cepa.org/article/information-chaos-plagues-serbia-elections/>.

Baletić, Katarina. "Serbia Reacts Fast Over AI Deepfake Video of PM – Unlike Other Cases." *Balkan Insight*, July 4, 2024. <https://balkaninsight.com/2024/07/04/serbia-reacts-fast-over-ai-deepfake-video-of-pm-unlike-other-cases/bi/>.

Barari, Soubhik, Christopher Lucas, and Kevin Munger. 2025. "Political Deepfakes Are as Credible as Other Fake Media and (Sometimes) Real Media." *The Journal of Politics* 87.2. 510-526.

Bertelsmann Stiftung. BTI 2026 Country Report: Serbia. Gütersloh: Bertelsmann Stiftung, 2026. <https://bti-project.org/en/reports/country-report/SRB>.

Beta. "Vučić o navodnom snimku Zobenice: Poslao mi je poruku da je to veštačka inteligencija." *N1*, November 28, 2024. <https://n1info.rs/vesti/vucic-o-navodnom-snimku-zobenice-poslao-mi-je-poruku-da-je-to-vestacka-inteligencija/>.

Bezbradica, Lazar, Ana Kosanović, Borjana Georgiou Sekuloski, and Ratko Stajić. 2026. "AI Deepfake Technologies." *Serbian Journal of Engineering Management* 11, Special Issue. 54–63.

Bieber, Florian. 2018. "Patterns of Competitive Authoritarianism in the Western Balkans." *East European Politics* 34.3. 337–354.

Bode, Ingvild. 2024. "AI Technologies and International Relations: Do We Need New Analytical Frameworks?" *The RUSI Journal* 169.5. 66-74.

Bokić, Uglješa. "Odbačena krivična prijava protiv Zobenice zbog audio snimka: Tužilaštvo zaključilo da nema elemenata krivičnog dela." *Danas*, February 25, 2025. <https://www.danas.rs/vesti/drustvo/odbacena-krivicna-prijava-protiv-zobenice-zbog-audio-snimka-tuzilastvo-zakljucilo-da-nema-elemenata-krivicnog-dela/>.

Bošnjaković, Željko. "'Vrlo opasna manipulacija': Sagovornici Danasa o fabrikovanim izjavama Dilasa, Jeremića i Marinike Tepić u režiji Željka Mitrovića." *Danas*, August 15, 2023. <https://www.danas.rs/vesti/drustvo/zeljko-mitrovic-dragan-djilas-vestacka-inteligencija-vuk-jeremic-marinika-tepic/>.

Bursać, Dejan, and Dušan Vučićević. 2025. "Electioneering Strategies in the Hybrid Regime: Another Regular Snap Election in Serbia." *New Perspectives* 33.1. 64–76.

Chesney, Bobby, and Danielle Citron. 2019. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." *California Law Review* 107.6. 1753–1820.

Ćirić, Sonja. "Arhiv: U Srbiji se protestuje u 400 mesta." *Vreme*, March 9, 2025. <https://vreme.com/vesti/arhiv-u-srbiji-se-protestuje-u-400-mesta/>.

CMS. "AI Laws and Regulations in Serbia." CMS Expert Guide, February 17, 2026. <https://cms.law/en/int/expert-guides/ai-regulation-scanner/serbia>.

CRTA. *Kampanja pre raspisivanja: Izveštaj Posmatračke misije Crte – izborna godina 2026/2027*. Belgrade: CRTA.

CRTA. Media Monitoring 2025. Belgrade: CRTA, 2026. https://crt.rs/wp-content/uploads/2026/01/Media-Monitoring_2025_ENG.pdf.

De Mesquita, E. Bueno, et al. 2023. "Preparing for generative AI in the 2024 election: Recommendations and best practices based on academic research." *University of Chicago Harris School of Public Policy and the Stanford Graduate School of Business*.

De Nadal, Lluís, and Peter Jančárik. 2024. "Beyond the deepfake hype: AI, democracy, and "the Slovak case".". *HKS Misinformation Review* 5.4.

Đokić, Danica. "Pink u vestima prikazivao lažne izjave Save Manojlovića kreirane uz pomoć veštačke inteligencije. REM ne reaguje." *Cenzolovka*, May 17, 2024. <https://www.cenzolovka.rs/drzava-i-mediji/pink-u-vestima-prikazivao-lazne-izjave-save-manojlovica-kreirane-uz-pomoc-vestacke-inteligencije-rem-ne-reaguje/>.

Đokić, Danica. "REM pokrenuo postupak protiv Pinka zbog snimka Save Manojlovića montiranog uz pomoć veštačke inteligencije." *Cenzolovka*, May 28, 2024. <https://www.cenzolovka.rs/drzava-i-mediji/rem-pokrenuo-postupak-protiv-pinka-zbog-snimka-sava-manojlovica-montiranog-iz-pomoc-vestacke-inteligencije/>.

Đorđević, Tijana. "TV Pink's New Shows – What Are the Limits of Parody?" *Istinomer*, August 17, 2023. <https://english.istinomer.rs/analyses/tv-pinks-new-shows-what-are-the-limits-of-parody/>.

Dragojlov, Aleksandra. 2026. "Influence of Political Clientelism on Media Freedom under Vucic and the Progressive Party." *Nationalities Papers* 54.2. 391–407.

Duberry, Jérôme. 2022. *Artificial Intelligence and Democracy: Risks and Promises of AI-Mediated Citizen–Government Relations*. Cheltenham, UK: Edward Elgar Publishing.

Džunić, Marija, and Nataša Golubović. 2026. "Regenerating Democracy in a Captured Media Environment: Student Protests and Hypermedia Mobilization in Serbia." *Frontiers in Political Science* 8.

Ejdus, Filip, and Tijana Rečević. 2021. "Ontological insecurity as an emergent phenomenon." *European Psychologist*.

Ejdus, Filip. 2018. "Critical Situations, Fundamental Questions and Ontological Insecurity in World Politics." *Journal of International Relations and Development* 21.4. 883–908.

Ejdus, Filip. 2020. *Crisis and Ontological Insecurity: Serbia's Anxiety over Kosovo's Secession*. Cham: Palgrave Macmillan.

European Federation of Journalists et al. A Spiral of Violence against Journalists: Serbia's Enduring Media Freedom Crisis. Brussels: Media Freedom Rapid Response and Council of Europe Safety of Journalists Platform, 2026. <https://europeanjournalists.org/blog/2026/07/16/serbia-a-dangerous-spiral-of-violence-against-journalists/>.

Fischer, Renate, and Otfried Jarren. 2024. "The platformization of the public sphere and its challenge to democracy." *Philosophy & Social Criticism* 50.1. 200-215.

FoNet. "Savet REM-a se oglasio o veštačkoj inteligenciji na Pinku." N1, August 21, 2023. <https://n1info.rs/vesti/savet-rem-a-se-oglasio-o-vestackoj-inteligenciji-na-pinku/>.

Freedom House. Freedom in the World 2026: Serbia. Washington, DC: Freedom House, 2026. <https://freedomhouse.org/country/serbia/freedom-world/2026>.

Freedom House. Nations in Transit 2024: Serbia. (2024). <https://freedomhouse.org/country/serbia/nations-transit/2024>.

Garbe, Lisa, and Seraphine F. Maerz. 2026. "The rise of authoritarian informationalism: escalating surveillance, manipulation, and control." *Democratization* 33.1. 1-18.

Gercama, Ingrid. "'We Are Done with Corruption': How the Students of Serbia Rose Up against the System." The Guardian, January 30, 2025. <https://www.theguardian.com/world/2025/jan/30/we-are-done-with-corruption-how-the-students-of-serbia-rose-up-against-the-system>.

Giddens, Anthony. 1991. *Modernity and Self-Identity: Self and Society in the Late Modern Age*. Cambridge: Polity Press.

Giebel, Katja. "Will Serbia's Protest Movement Lead to a Democratic Beginning?" Heinrich Böll Stiftung, March 13, 2025. <https://www.boell.de/en/2025/03/13/will-serbias-protest-movement-lead-to-a-democratic-beginning>.

Government of the Republic of Serbia. *Strategy for the Development of Artificial Intelligence in the Republic of Serbia for the Period 2025–2030*. Official Gazette of the Republic of Serbia, no. 5/2025.

Haas, Peter M. 1992. "Introduction: Epistemic Communities and International Policy Coordination." *International Organization* 46.1. 1–35.

Habgood-Coote, Joshua. 2023. "Deepfakes and the Epistemic Apocalypse". *Synthese* 201.3. Article 103.

Horowitz, Michael C. 2018. "Artificial Intelligence, International Competition, and the Balance of Power." *Texas National Security Review* 1.3. 37-57.

Jovović, Jelena, and Stefan Janjić. Osvrt na dezinformacije u 2025. godini. Novi Sad: Novosadska novinarska škola, 2026. <https://fakenews.rs/wp-content/uploads/2025-Unatrag.pdf>.

Jungherr, Andreas. 2023. *Artificial Intelligence and Democracy: A Conceptual Framework*. Social Media + Society 9.3.

Karaica, Sanela. 2025. "Unreliable AI Tools for Detecting Fake Content." MC_ONLINE, July 30.

Kinnvall, C. 2006. "Globalization and religious nationalism in India: The search for ontological security". *Globalization and Religious Nationalism in India: The Search for Ontological Security*, 25(5).

Kofi Annan Foundation. Electoral Vulnerability Index 2026–2027: Country Analysis-Serbia. Geneva: Kofi Annan Foundation, 2026. <https://www.kofiannanfoundation.org/wp-content/uploads/2026/06/Serbia-Country-Analysis-Electoral-Vulnerability-Index-2026-2027.pdf>.

Kosanović, Stefan. "Stručnjaci za zvuk: Snimak Damira Zobenice gotovo sigurno autentičan." Raskrikavanje, November 29, 2024. <https://raskrikavanje.rs/strucnjaci-za-zvuk-snimak-damira-zobenice-gotovo-sigurno-autentican/>.

Kossev. "Video Published Targeting Ivan Ninić as a CIA Spy and Collaborator of the 'Albanian Occupier.'" KoSsev, August 24, 2025. <https://kossev.info/en/objavljen-spot-kojim-se-ivan-ninic-targetira-kao-spijun-cia-i-saradnik-albanskog-okupatora/>.

Krasteva, Anna, and Nebojša Vladislavljević. 2017. "Securitisation versus Citizenship in the Balkan States: Populist and Authoritarian Misuses of Security Threats and Civic Responses." *Global Campus Human Rights Journal* 1. 373-391.

Krstić, Aleksandra, Filip Ejduš, and Nebojša Vladislavljević. 2022. "The Politics of Giving Voice: Media Framing of Extremist Groups in the 2010 Pride Parade in Serbia." *Problems of Post-Communism*.

Krstić, Aleksandra. 2023. "Digital Transformation of Journalism and Media in Serbia: What Has Gone Wrong?" *Journalism* 25.5. 1014-1030.

Kurowska, Xymena. 2023. "Epistemic security and the redemptive hegemony of magical realism." *Cambridge Review of International Affairs*.

Łabuz, Mateusz, and Christopher Nehring. 2024. "On the Way to Deep Fake Democracy? Deep Fakes in Election Campaigns in 2024." *European Political Science* 23.4. 454-473.

Lazić, Miloš Ž. "Ivan Ninić o spotu u kojem ga opasno targetiraju: 'S druge strane nisu književnici, prodavci kokica i semenki već okoreli kriminalci koji brane ono što su pokrali.'" Nova.rs, August 24, 2025. <https://nova.rs/vesti/hronika/ivan-ninic-o-spotu-u-kojem-ga>

[opasno-targetiraju-s-druge-strane-nisu-knjizevnici-prodavci-kokica-i-semenki-vec-okoreli-kriminalci-koji-brane-ono-sto-su-pokrali/](#).

Levitsky, Steven, and Lucan A. Way. 2010. *Competitive Authoritarianism: Hybrid Regimes after the Cold War*. Cambridge: Cambridge University Press.

M. D. “AJMO, STOKO, DA RUŠITE ZA MENE’: Ovaj snimak Đilasa pokazuje kako je lako MANIPULISATI POMOĆU VEŠTAČKE INTELIGENCIJE (AUDIO).” Republika.rs, November 28, 2024, 20:44. <https://www.republika.rs/vesti/politika/605301/dragan-djilas-damir-zobenica-marinika-tepic>.

Maerz, Seraphine F. 2026. “How Practices of Digital Authoritarianism Harm Democracy.” *Democratization* 33.1. 163-190.

Marković, Ana. “Fotografije napada batinaša na studente i profesore FDU koji su odavali poštu poginulima u Novom Sadu.” Nova.rs, November 22, 2024. <https://nova.rs/vesti/drustvo/napad-besnih-vozaca-na-neduzne-studente-i-profesore-fdu/>.

Milivojević, Snježana. “Serbia.” Digital News Report 2025. Oxford: Reuters Institute for the Study of Journalism, University of Oxford, 2025. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2026/serbia>.

Milutinović, Irina, and Aleksandra Krstić. 2026. “Countering Disinformation Amid Democratic Backsliding: Difficult Comparisons Between the European Union and Serbia.” *Journalism and Media* 7.1.

Mitzen, Jennifer. 2006. “Ontological Security in World Politics: State Identity and the Security Dilemma.” *European Journal of International Relations* 12.3. 341-370.

N1 Belgrade. “Court Orders Pink TV to Stop Using AI-Generated Video of Djilas.” N1, August 14, 2024. <https://n1info.rs/english/news/court-orders-pink-tv-to-stop-using-ai-generated-video-of-djilas/>.

N1 Belgrade. “Prime Minister’s Office: Fake Video of Vucevic Posted on Facebook.” N1, July 3, 2024. <https://n1info.rs/english/news/prime-minister-s-office-fake-video-of-vucevic-posted-on-facebook/>.

N1 Beograd. “(AUDIO) Tepić: SNS funkcioner daje zadatke za blokadu – izazivanje incidenata i vređanje opozicije.” N1, November 28, 2024. <https://n1info.rs/vesti/audio-tepic-sns-funkcioner-daje-zadatke-za-blokadu-izazivanje-incidenata-i-vredjanje-opozicije/>.

N1 Beograd. “Draža Petrović: ‘Veštačka inteligencija’ Damira Zobenice izvela na ulice prirodnu mladu inteligenciju.” N1, December 27, 2024. <https://n1info.rs/vesti/draza-petrovic-vestacka-inteligencija-damira-zobenice-izvela-na-ulice-prirodnu-mladu-inteligenciju/>.

Nawrocki, Michał. 2024. “After the epistemological turn: A framework for studying populism as a knowledge phenomenon.” *The British Journal of Politics and International Relations* 26.2. 444-465.

Nikolić, Maja. “Možda je inteligencija veštačka, ali su objašnjenja glupa’: Ko je članovima SNS naredio da prave incidente.” N1, November 29, 2024. <https://n1info.rs/vesti/mozda-je-inteligencija-vestacka-ali-su-objasnjenja-glupa-ko-je-clanovima-sns-naredio-da-prave-incidente/>.

NIN. “Studenti o skupu od 15. marta i daljim planovima: Tu smo da dugoročno gradimo.” NIN, March 21, 2025. <https://www.nin.rs/politika/vesti/71569/studenti-o-skupu-od-15-marta-i-daljim-planovima-tu-smo-da-dugorocno-gradimo>.

Nord, Marina, David Altman, Fabio Angiolillo, Tiago Fernandes, Ana Good God, and Staffan I. Lindberg. 2025. *Democracy Report 2025: 25 Years of Autocratization – Democracy Trumped?* Gothenburg: V-Dem Institute, University of Gothenburg. https://www.v-dem.net/documents/54/v-dem_dr_2025_lowres_v1.pdf.

Nordicom. 2026. *Nordic News Media Landscapes 2025*. Gothenburg: Nordicom, University of Gothenburg.

Novaković, Ana. “REM i Željko Mitrović u epizodi veštačka inteligencija: ‘To je jedna farsa.’” N1, August 22, 2023. <https://n1info.rs/vesti/rem-i-zeljko-mitrovic-u-epizodi-vestacka-inteligencija-ili-manipulacija/>.

NUNS. “Vučić o zloupotrebi veštačke inteligencije na Pinku: To ne sme da se radi.” NUNS, August 18, 2023. <https://nuns.rs/vucic-o-zloupotrebi-vestacke-inteligencije-na-pinku-to-ne-sme-da-se-radi/>.

Office of the United Nations High Commissioner for Human Rights (OHCHR). “Serbian Authorities Need to Speed Up Steps to Ensure Accountability – UN Human Rights Chief.” November 5, 2025. <https://www.ohchr.org/en/press-releases/2025/11/serbian-authorities-need-speed-steps-ensure-accountability-un-human-rights>.

OSCE Office for Democratic Institutions and Human Rights (ODIHR). Serbia, Early Parliamentary Elections, 17 December 2023: Final Report. Warsaw: OSCE/ODIHR, 2024. <https://odihr.osce.org/odihr/elections/serbia/563502>.

Parliamentary Assembly of the Council of Europe (PACE). 2024. Observation of the Early Parliamentary Elections in Serbia (17 December 2023). Election Observation Report, Doc. 15895. Strasbourg: Council of Europe. <https://pace.coe.int/en/files/33240/html>.

Pawelec, Maria. 2022. “Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions”. *Digital Society* 1.2.

Pew Research Center. 2025. *How People Around the World View AI*. Washington, DC: Pew Research Center. <https://www.pewresearch.org/global/2025/10/15/how-people-around-the-world-view-ai/>.

Predić, Ivana. “Šta to REM krije od građana: Objašnjenja odluka nema, zapisnici sve skromniji, informacije i dokumenti nedostupniji.” Cenzolovka, July 16, 2024.

<https://www.cenzolovka.rs/drzava-i-mediji/sta-to-rem-krije-od-gradjana-objasnjenja-odluka-nema-zapisnici-sve-skromniji-informacije-i-dokumenti-nedostupniji/>.

ProGlas. "ProGlas objavio detalje akcije 'Zastani Srbijo!' zakazane za petak." November 20, 2024. <https://proglas.co.rs/2024/11/20/proglas-objavio-detalle-akcije-zastani-srbijo-zakazane-za-petak/>.

Radio Slobodna Evropa. "REM o veštačkoj inteligenciji: Obaveza medija da spreče zloupotrebu nečijeg lika i glasa." Radio Slobodna Evropa, August 21, 2023. <https://www.slobodnaevropa.org/a/vestacka-inteligencija-rem-pink/32557422.html>.

Radio Televizija Pančevo. "Ekskluzivno! Imamo priznanje gospođe koja je dobila pare od Đilasove pudlice Maje Dražić!" RTV Pančevo, January 31, 2025. <https://rtvpancevo.rs/vest/4458/ekskluzivno-imamo-priznanje-gospode-koja-je-dobila-pare-od-dilasove-pudlice-maje-drazic>.

Reci, Megi. "Digital Rights Review: AI and Growing Uncertainty Over What's Real Online." Balkan Insight, September 25, 2025. <https://balkaninsight.com/2025/09/25/digital-rights-review-ai-and-growing-uncertainty-over-whats-real-online/bi/>.

Rini, Regina. 2020. "Deepfakes and the Epistemic Backstop." *Philosophers' Imprint* 20.24. 1–16.

Seger, Elizabeth, Shahar Avin, Gavin Pearson, Mark Briers, Seán Ó Heigeartaigh, and Helena Bacon. 2020. *Tackling Threats to Informed Decision-Making in Democratic Societies: Promoting Epistemic Security in a Technologically-Advanced World*. London: The Alan Turing Institute.

Simon, Felix M., Rasmus Kleis Nielsen, and Richard Fletcher. 2025. *Generative AI and News Report 2025: How People Think about AI's Role in Journalism and Society*. Oxford: Reuters Institute for the Study of Journalism.

Simon, Julia. 2024. "Problematizing modern democracy in the United States: An 'insurrection of subjugated knowledges' in the wake of the 2020 presidential election." *Politische Vierteljahresschrift* 65.2. 367-394.

Srećković, Aleksandra, Vladana Jaraković, Marija Stefanović, Tamara Branković, Lana Avakumović, and Milena Manojlović. 2022. *Mapping the Media Landscape in Serbia 2020–2021*. Belgrade: CRTA. <https://crtars/en/mapping-the-media-landscape-in-serbia-2020-2021/>.

Sremskomitrovački portal. "ŠOK SNIMAK! Đilasov naredba osvanula na društvenim mrežama: 'Trebaju nam incidenti u Rumi.'" January 31, 2025. <https://mitrovica.info/sok-snimak-djilasov-naredba-osvanula-na-drustvenim-mrezama-trebaju-nam-incidenti-u-rumi/>.

Steele, Brent J. 2008. *Ontological security in international relations*. New York: Routledge.

Student Blockades. "About the Blockades." <https://blokada.org/o-nama/>.

Subotić, Jelena. 2016. "Narrative, ontological security, and foreign policy change." *Foreign policy analysis* 12.4. 610-627.

Twomey, John, Didier Ching, Matthew Peter Aylett, Michael Quayle, Conor Linehan, and Gillian Murphy. 2023. *Do Deepfake Videos Undermine Our Epistemic Trust? A Thematic Analysis of Tweets That Discuss Deepfakes in the Russian Invasion of Ukraine*. PLOS ONE 18.10.

Vaccari, Cristian, and Andrew Chadwick. 2020. *Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News*. Social Media + Society 6.1.

Vladisavljević, Nebojša, and Aleksandra Krstić. 2023. "Competitive Authoritarianism and Populism in Serbia under Vučić in Political Cartoons." *Europe-Asia Studies* 75.9. 1429-1452.

Vladisavljević, Nebojša. 2020. "Media Discourse and the Quality of Democracy in Serbia after Milošević." *Europe-Asia Studies* 72.1. 8-32.

Vorlíček, Dagmar. 2026. "Between conspiracy and critique: the politics of epistemic insecurity in the "lab leak controversy"." *Security Dialogue* 57.1. 84-96.

Vreme. "Na Fejsbuku objavljena lažna izjava Vučevića – Više javno tužilaštvo pokrenulo istragu." Vreme, July 4, 2024. <https://vreme.com/en/vesti/na-fejsbuku-objavljena-lazna-izjava-vucevica-vise-javno-tuzilastvo-pokrenulo-istragu/>.

West Health and Gallup. 2026. *Millions of Americans Now Consult AI Before, After, and Sometimes Instead of, Seeing a Doctor*. Washington, DC: West Health–Gallup Center on Healthcare in America.

Westerlund, Mika. 2019. "The Emergence of Deepfake Technology: A Review." *Technology Innovation Management Review* 9.11.39-52.

Yin, Robert K. 2018. *Case Study Research and Applications: Design and Methods*. 6th ed. Thousand Oaks, CA: SAGE Publications.

Zarakol, Ayşe. 2010. "Ontological (in) security and state denial of historical crimes: Turkey and Japan." *International relations* 24.1. 3-23.

Živić, Miona. 2024. "Veštačka inteligencija ili Zobenica 'stasom i glasom': Stručnjak tvrdi – nemoguće da AI napravi tako autentičan snimak." N1, November 29, 2024. <https://n1info.rs/vesti/vestacka-inteligencija-ili-zobenica-stasom-i-glasom-strucnjak-tvrdi-nemoguće-da-ai-napravi-tako-autentican-snimak/>.