



# EPHEMERIDES

---

1-2/2025

**STUDIA  
UNIVERSITATIS BABEŞ-BOLYAI  
EPHEMERIDES**



**1-2/2025  
January - December**

## EDITORIAL BOARD

### EXECUTIVE EDITORS:

Professor PhD ANDREEA MOGOȘ, Babeș-Bolyai University

Lecturer PhD SINZIANA JURĂU, Babeș-Bolyai University

Assoc. Professor PhD RADU MEZA, Babeș-Bolyai University

### EDITORIAL BOARD:

Professor ILIE RAD, Ph.D., Babeș-Bolyai University Cluj-Napoca, Romania

Professor CSEKE PETER, Ph.D., Babeș-Bolyai University Cluj-Napoca, Romania

Professor AUREL CODOBAN, Ph.D., Dimitrie Cantemir Christian University, Cluj-Napoca, Romania

Professor MIHAI COMAN, Ph.D., University Bucharest, Romania

Associate Professor GEORGETA DRULA, Ph.D., University of Bucharest, Romania

Associate Professor MARIANA CERNICOVA-BUCA, Ph.D., Polytechnic University of Timisoara, Romania

Associate Professor RAMONA HOSU, Ph.D., Babeș-Bolyai University Cluj-Napoca, Romania

Assistant Professor SÎNZIANA JURĂU, Ph.D., Babeș-Bolyai University Cluj-Napoca, Romania

Associate Professor ANDREEA MOGOȘ, PhD, Babeș-Bolyai University Cluj-Napoca, Romania

Associate Professor CRISTINA NISTOR, PhD, Babeș-Bolyai University Cluj-Napoca, Romania

Associate Professor ZOLTÁN TIBORI SZABÓ, PhD, Babeș-Bolyai University Cluj-Napoca, Romania

Professor SABINA ADDAMIANO, PhD, Roma Tre University, Italy

Professor MALI ALON, PhD, Tiltan School of Design and Visual Communication, Haifa, Israel

Professor CHARO LACALLE ZALDUENDO, PhD, Autonomous University of Barcelona, Spain

Professor JOHN HARTSOCK, PhD, State University of New York College at Cortland, USA

Professor KENETH STARCK, PhD, School of Journalism and Communication, Iowa, USA

Professor NILÜFER PEMBECIOĞLU, PhD, State University, Istanbul, Turkey

Professor PETER GROSS, PhD, University of Tennessee Knoxville, USA

Professor TUDOR VLAD, PhD, James M. Cox Jr. Center for International Mass Communication Training and Research, University of Georgia, USA

**Studia Universitatis Babeș-Bolyai Ephemerides, [studia.ephemerides@fspac.ro](mailto:studia.ephemerides@fspac.ro)**

© Studia UBB Ephemerides. Published by Babeș-Bolyai University

ISSN (online): 2065-9555 | ISSN (print): 1224-872X | ISSN-L: 2065-9555

Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

**YEAR**  
**MONTH**  
**ISSUE**

**Volume 70 (LXX) 2025**  
**DECEMBER**  
**1-2**

---

PUBLISHED ONLINE: 2025-12-30

PUBLISHED PRINT: 2025-12-30

ISSUE DOI:10.24193/subbeph.2025

---

**S T U D I A**  
**UNIVERSITATIS BABEŞ-BOLYAI**  
**EPHEMERIDES**

**1-2**

**ISSUE EDITOR:**

Andreea MOGOŞ, Babeş-Bolyai University Cluj-Napoca, Romania

**C O N T E N T**

- Luciana RĂDUŢ-GAGHI, Adel TAYEBI, *L'autonomie stratégique européenne entre mode discursive et besoin géopolitique / European Strategic Autonomy: Between Discursive Trend and Geopolitical Necessity* ..... 5
- Francesca BISIANI, *Ce qu'on peut dire de l'autre : dynamiques de légitimation des discours antiféministes / What We Can Say About the Other: The Dynamics of Legitimation in Anti-feminist Discourse* ..... 35
- Jérôme ROUDIER, *La construction problématique de l'« Italie » chez Machiavel, une modélisation déjà moderne ? / The Problematic Construction of 'Italy' in Machiavelli: An Already Modern Model?* ..... 53

|                                                                                                                                                                                                                      |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Cristina LÓPEZ GONZÁLEZ, Carolina BEIRO GARCIA, Nuria FIGUEIRA MARQUEZ, <i>Media Discourse and Political Polarization in the Coverage of the Valencia DANA. A Comparative Analysis of Público and Okdiario</i> ..... | 65  |
| Erika-Zita CSÓKA-ANTAL, <i>Who Are We in the World of GIFs? – Mapping GIF Users</i> .....                                                                                                                            | 81  |
| Georgiana Silvia LEOTESCU, <i>Charting Ethical Communication in Relation to Integrity in Scientific Research</i> .....                                                                                               | 97  |
| Ramona-Gabriela HOSU, Sînziana JURĂU, <i>Health and Identity in between the Private and the Institutional. The Case of Catherine, the Princess of Wales</i> .....                                                    | 113 |
| Mirela ABRUDAN, Carmen Ioana NEAMȚU, <i>Deepfakes, Fake News, And Digital Trust: Insights from Romanian Social Media Users</i> .....                                                                                 | 145 |
| Liu QINGTAO, <i>The Dual Construction of Generative Artificial Intelligence Domestication Practices and Information Cocoons. An Empirical Study Based on College Students' Usage Patterns</i> .....                  | 167 |

## L'autonomie stratégique européenne entre mode discursive et besoin géopolitique

Luciana RĂDUȚ-GAGHI<sup>1</sup>, Adel TAYEBI<sup>2</sup>

**Abstract. European Strategic Autonomy: Between Discursive Trend and Geopolitical Necessity.** This article examines how the concept of European strategic autonomy circulates and is appropriated across institutional and media discourses in the French-speaking public sphere. Drawing on discourse analysis and the framework of *discursive appropriation*, the authors investigate whether strategic autonomy is summoned by current events or represents a distinct discursive moment in European political communication. The study proceeds in two stages: first, a genealogical analysis of strategic autonomy's presence in official European Union documents since its emergence in the 1994 French White Paper on Defense, and second, a computational discourse analysis applied to a corpus of French-language newsmedia articles. The research reveals that while EU institutions employ the concept in limited and relatively sparse ways, strategic autonomy has become increasingly salient in media discourse and public debate, particularly following renewed geopolitical tensions. The article argues that strategic autonomy functions as a *metaphorical concept* that operates simultaneously at political and metapolitical levels, serving to reconfigure European identities, institutional relationships, and the boundaries of public debate. By comparing institutional and media appropriations, the study demonstrates how different discursive instances cooperate in stabilizing or destabilizing the meaning of strategic autonomy within the broader European public space.

**Keywords:** Strategic autonomy, European institutions, media discourse, discursive appropriation

---

<sup>1</sup> Professeur des universités, CY Cergy Paris Université, laboratoire LT2D, luciana.radut-gaghi@cyu.fr

<sup>2</sup> Docteur EUTOPIA, CY Cergy Paris Université (Laboratoire LT2D) et University Nova Lisboa, adel.tayebi@cyu.fr



Il existe des concepts à la fois inclusifs et exclusifs, programmatiques et législatifs, opérationnels et stratégiques. L'« autonomie stratégique » en est un. Il traite de l'identité et de l'altérité en égale mesure. Associé à l'Europe depuis le Livre blanc sur la défense de 1994, il est surtout connu depuis la Stratégie globale de l'Union européenne de 2016. Depuis la seconde accession au pouvoir de Donald Trump, l'autonomie stratégique européenne est plus que jamais à l'ordre du jour des politiques, des institutions et des médias.

Dans une perspective d'analyse du discours, dans cet article, nous analysons la circulation du concept d'autonomie stratégique dans l'espace discursif francophone. La notion d'appropriation discursive (Radut-Gaghi, 2021) nous est utile ici car elle rend compte à la fois de la manière dont l'instance discursive reçoit des idées et des discours et la manière dont elle imprègne son propre discours de valeurs, connaissances, aspirations. Cette approche nous permet de rendre compte du dynamisme et de la conflictualité de la circulation du concept d'autonomie stratégique (Amossy, 2014 ; Chilton, 2004 ; Mouffe, 2005) ainsi que de son caractère « métapolitique » (Zinkowski, 2019).

Nous formulons cet article autour d'une question de recherche principale : Est-ce l'actualité qui sollicite l'autonomie stratégique par moments, ou assiste-t-on à moment discursif de l'autonomie stratégique ? En fin de compte, en quelle mesure les différentes instances discursives coopèrent dans la stabilisation du sens de l'autonomie stratégique ?

Pour ce faire, nous procédons en deux volets. Premièrement, nous réalisons une généalogie de la présence thématique de l'autonomie stratégique dans les actes officiels de l'Union européenne selon des procédés identifiés par Oger et Olliver-Yanniv (2003) et Krieg Planque et Oger (2012). Deuxièmement, nous nous appuyons sur un corpus de presse auquel nous appliquons une analyse computationnelle fondée sur un emploi réflexif de LLM afin de rendre compte des caractéristiques générales de ce corpus, des polarités et des phénomènes d'identification/opposition. Une analyse des arguments complète cette inspection du corpus médiatique. Une partie finale se propose de comparer les résultats des deux volets et de discuter la place de l'autonomie stratégique à l'échelle de l'espace public européen aujourd'hui.

## **1. Une approche conceptuelle : Les appropriations discursives**

Dans un espace de la parole publique libre, les idées, les représentations et les arguments circulent entre différentes instances discursives et se transforment par la même occasion, en fonction des intérêts, des pressions, des objectifs

des participants au débat public. Dans le cadre de nos travaux précédents (Radut-Gaghi *et al*, 2017 ; Radut-Gaghi, 2021), nous identifions ce jeu de la circulation et de l'ajustement des idées publiques comme des procédures d'appropriation discursives. L'appropriation serait donc une opération de contrôle, d'organisation du discours selon l'analyse de l'ordre du discours foucauldienne (1971).

Le concept était déjà présent dans la littérature des humanités. Dans une sémiotique des pratiques interprétatives, l'appropriation intervient après la compréhension et la familiarisation, qui elle-même implique l'appréciation : « l'appropriation est à la base des conditions de diffusion et de pérennisation d'une culture mais elle peut favoriser aussi la résistance de l'altérité à l'homogénéisation conditionnée par l'empire d'une épistémè, avec tous ses plis autoréflexifs » (Basso Fossali, 2018, 13). L'intérêt de cette approche est de rendre compte de son ambivalence, de sa bilatéralité : « La bilatéralité de l'appropriation (s'approprier et se rendre propre à) déplace la prééminence de l'initiative du sujet au profit de la reconnaissance d'un caractère presque événementiel de l'appropriation même. » (ibid, 21)

Cette approche vient conforter notre hypothèse selon laquelle les processus d'appropriation opèrent à trois niveaux distincts et articulés : le niveau individuel, celui de l'auteur en tant que sujet du discours ; le niveau médiatique, correspondant à l'instance ou au dispositif de communication ; et le niveau institutionnel, entendu comme l'ensemble des valeurs, des codes et des pratiques qui le structurent (Radut-Gaghi, 2021).

À l'échelle institutionnelle, l'appropriation se manifeste par un effacement relatif de l'acteur au profit des contraintes propres à l'institution, lesquelles imposent leur manière de définir et de traiter l'objet du discours. Toutefois, des travaux issus des sciences de l'information et de la communication, tout comme des sciences politiques, s'attachent à réhabiliter la part de l'acteur dans la mobilisation des idées au sein des cadres institutionnels.

Au niveau médiatique, l'appropriation renvoie à la façon dont l'instance médiatique se saisit d'une représentation et en organise la diffusion auprès du public. Il ne s'agit donc pas encore de l'énonciation au sens strict, mais bien du dispositif médiatique. Le discours médiatique apparaît ainsi comme un ensemble structuré de possibilités techniques, de choix de hiérarchisation et de modalités de mise en contenu.

Enfin, au niveau individuel, l'appropriation relève d'un processus étroitement lié à la subjectivité de l'acteur : elle dépend à la fois de ses conditions psychosociologiques, de ses valeurs, opinions et croyances, mais aussi de

processus cognitifs tels que la rationalité et les mécanismes de sélection qu'il opère.

Afin d'inscrire les appropriations dans une conception élargie de la politique, un détour par la *Critical Discourse Analysis* s'avère pertinent. Paul Chilton (2004) propose une approche de la politique articulée autour de deux dimensions : d'une part, la lutte pour le pouvoir et, d'autre part, la pratique coopérative de résolution des conflits. Chacune de ces dimensions se déploie à deux niveaux. Le niveau micro renvoie aux conflits d'intérêts, aux luttes pour la domination, mais aussi aux formes de coopération, tandis que le niveau macro concerne les institutions politiques de l'État, envisagées à la fois comme des lieux de domination de certains groupes et comme des espaces de résolution des conflits d'intérêts.

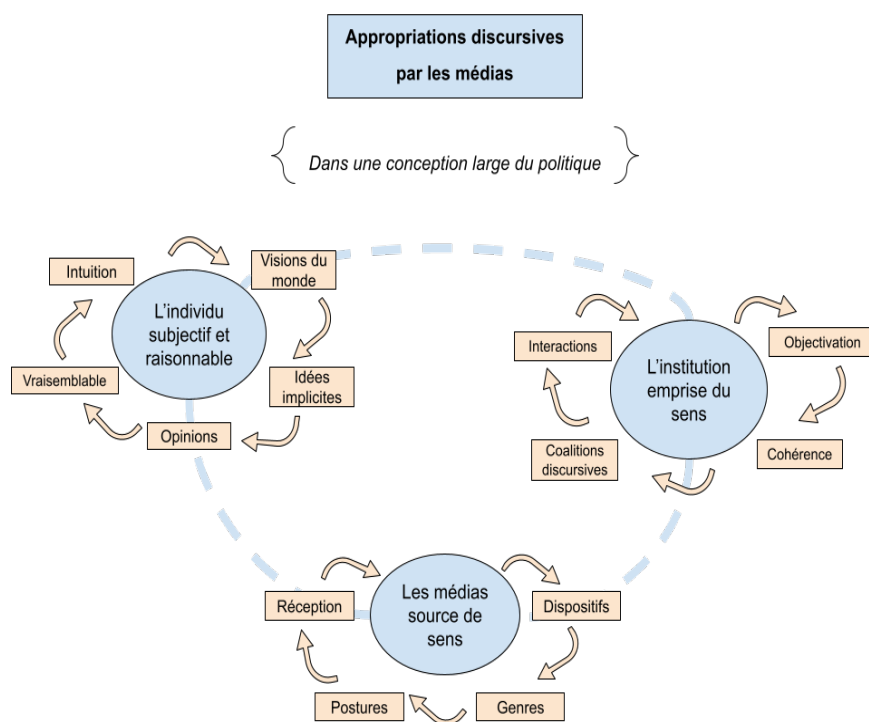
Une perspective encore plus englobante est développée par David Machin et Theo van Leeuwen (2016), qui appréhendent la politique à un niveau très large, diffusée à travers le discours culturel médiatisé. Celui-ci se manifeste aussi bien dans les médias traditionnels et numériques que dans le design, l'architecture, la mode ou, plus largement, dans tout dispositif culturel.

Cette conception élargie de la politique trouve un prolongement théorique dans les travaux de Chantal Mouffe (2005), qui distingue la sphère institutionnelle de la décision politique d'un domaine extra-parlementaire où se jouent les rapports de pouvoir. Elle oppose ainsi « la politique », entendue comme l'ensemble des pratiques et des institutions par lesquelles un ordre est établi et la coexistence humaine organisée, au « politique », qu'elle définit comme la dimension d'antagonisme constitutive des sociétés humaines.

Ce « politique », envisagé dans son acception la plus englobante, est qualifié de « métapolitique » par Jan Zienkowski (2019). Si le débat politique porte initialement sur des enjeux concrets tels que le réchauffement climatique, les migrations ou la santé, il acquiert une dimension métapolitique dès lors que ses effets sont susceptibles de reconfigurer les relations sociales et les identités qui structurent l'espace public. Plus largement, les débats métapolitiques peuvent redéfinir les frontières entre les différentes sphères d'action et d'expertise — celles des citoyens, des organisations de la société civile, des partis politiques et des institutions de l'État.

Dans cette perspective, les appropriations apparaissent comme des pratiques situées au croisement du politique et du métapolitique. En investissant des dispositifs discursifs et culturels qui excèdent les cadres institutionnels

de la décision, elles participent à la reconfiguration des rapports de pouvoir, des identités et des frontières du débat public. Loin de se limiter à de simples usages ou détournements, les appropriations constituent ainsi des modes d'intervention politique à part entière, inscrits dans une conception élargie et discursive de la politique. La figure 1 propose une cartographie programmatique des appropriations discursives.



**Figure 1.** Un modèle des appropriations discursives (Radut-Gaghi, 2021)

Concernant l'appropriation de l'idée d'autonomie stratégique, qui retient notre intérêt dans cet article, nous proposons dans les pages qui suivent de l'appliquer aspects médiatiques et aux aspects institutionnels, atteignables à travers les discours publics. Chaque sous partie débute avec une prospection méthodologique et continue avec des résultats de la recherche.

## **2. Appropriations par les institutions**

Nous choisissons de dérouler ce volet de la recherche au niveau des institutions européennes et donc des discours qu'elles produisent. Bien que le concept d'autonomie stratégique ait une origine française, il est aujourd'hui largement appliqué à l'Union européenne.

### **2.1. Outils méthodologiques**

Claire Oger et Caroline Ollivier-Yaniv posent les types de définition des discours institutionnels selon des « cercles » ayant au cœur la définition juridique : « Nous entendons par 'discours institutionnel', au sens strict, le discours produit officiellement par un énonciateur singulier ou collectif qui occupe une position juridiquement inscrite dans l'appareil d'État, qu'il soit fonctionnaire ou représentant politique. » (Oger, Ollivier-Yaniv, 2003, 127). C'est à cette définition que nous ferons référence tout au long de cet article. Les deux autres cercles sont les discours produits par les politiques en dehors des contextes officiels, les cadres très souples comme la retraite, les essais, et enfin les « discours autorisés » mais sans référence à l'État (chartes, programmes, règlements...).

Selon une approche par la sociologie compréhensive, les deux auteurs conseillent d'approcher les discours du premier cercle par le « processus de consolidation » (Lacroix, Lagroye, 1992 – livre sur la Présidence française et étude sur les processus de consolidation des acteurs et de l'institution par des usages, des compétitions permises). Au niveau discursif, cela peut être difficile à saisir seulement en se fondant sur des documents écrits, comme nous le faisons ici. Mais effectivement, les entretiens, discours, échanges plus ou moins formels autour de la question de l'autonomie stratégique pourrait être fertile du point de vue de la sociologie compréhensive.

La « mise en scène discursive des institutions » n'est pas énormément exploitée par les auteurs francophones. Pourtant, le potentiel de connaissance est important : l'analyse du discours « ne se contente pas d'étudier des textes : elle les rapporte au fonctionnement des institutions qui les produisent et les gèrent », écrivait Dominique Maingueneau (2021, 58). Isabelle Laborde-Milaa en propose une analyse dans le recueil que Johannes Angermüller et Gille Philippe ont réalisé en hommage à D. Maingueneau. Cette auteure sollicite la notion de hyperénonciateur. « Dans le cas d'une institution, il s'agit d'une instance tout aussi complexe et plurielle [que la presse], mais constituée

d'emblée en tant que corps unifié, groupe où se fondent les individus, pourvu d'un nom et d'un logo souvent âprement trouvés et défendus » (Laborde-Milaa, 2015, 166). Les stratégies d'objectivation, le « nous » comme identité de la communauté, le « je » du locuteur authentique ou de la figure fictive et stéréotypique, la scénographie, etc., sont analysés comme autant de procédés de mise en discours des institutions.

L'analyse du discours appliquée aux institutions vise de manière prioritaire « la question de la légitimation des institutions » (2012). Alice Krieg-Planque et Claire Oger identifient ici deux volets dans l'analyse des discours institutionnels, qui mènent à des pistes de recherche distinctes. Le premier est lié à la « stabilisation des énoncés », sous forme de formules toutes faites ou de formules routinières. Le second est celui de « l'effacement de la conflictualité » : « réduire la dissonance, produire la cohérence », qui entraîne tout un programme de recherche - « la neutralisation de l'opposition entre posture d'expert et positionnement politique, la conciliation des intérêts entre partenaires publics et privés, la dénégaration des divergences d'opinion, ou encore la légitimation du discours savant » (ibid.). L'ouverture à ces pistes d'exploration rend compte de l'intérêt que peut avoir l'étude du discours des institutions dans la compréhension de la circulation des informations, des idées. Le crible institutionnel est révélateur pour le façonnage d'idées qui peuvent sembler par ailleurs comme allant de soi, et sont considérées comme acquises dans l'espace public élargi d'aujourd'hui.

Cette attention portée aux pratiques discursives et métapolitiques trouve un écho direct dans les travaux réunis dans l'ouvrage *Institutionality. Studies of Discursive and Material (Re-)ordering* (Scholz, Porsché, Singh, 2022). Les auteurs y mettent l'accent sur les effets de pouvoir produits par la construction de l'ordre institutionnel, à partir d'un ensemble d'études de terrain qui analysent les institutions non comme des cadres fixes, mais comme des configurations en constante (re)élaboration. La notion d'« institutionality » est ainsi mobilisée pour souligner le caractère construit de l'ordre institutionnel et pour inviter à l'analyser empiriquement comme un phénomène à la fois social, linguistique et matériel, générateur d'effets de pouvoir multiples et différenciés (ibid., 528). Dans cette perspective, les appropriations institutionnelles apparaissent comme des moments clés de cette (re)construction : en s'appropriant, en réinterprétant ou en réorganisant les dispositifs discursifs et matériels des institutions, les acteurs contribuent à produire, stabiliser ou transformer l'ordre institutionnel lui-même.

## ***2.2. Une généalogie du point de vue des institutions européennes***

Nous ne faisons ici qu'un historique synthétique du concept d'autonomie stratégique, dont la première occurrence dans un document officiel date de 1994 (Livre blanc sur la défense)<sup>3</sup>. Dans un contexte post-guerre froide habité par des interrogations sur le rôle et les garanties de l'OTAN, Marceau Long, Edouard Balladur et François Léotard ce document rédigé sous l'égide du premier ministre français évoque déjà l'autonomie stratégique (à la fois française et européenne) pour « la prévision et l'appréciation des conflits » (p. 78), par la participation « aux équilibres militaires en Europe et [la disposition] d'une capacité d'action sur la scène internationale » (p. 139).

Les aspects européens sont consolidés dans les conclusions du Conseil européen des 19 et 20 décembre 2013, quand l'accent est mis sur « la base industrielle et technologique de défense (BITDE) plus intégrée, plus durable, plu innovante et plus compétitive pour pouvoir assurer le développement et le soutien de ses capacités de défense » (p. 7). Il en est question une seule fois, tout de même.

L'inscription précise de l'autonomie stratégique dans les actes de l'Union européenne se fait en 2016 dans la « stratégie globale pour la politique étrangère et de sécurité de l'Union européenne » du Conseil de l'Union européenne.

L'autonomie stratégique n'est jamais tout à fait définie. Elle est présentée comme un besoin, comme étant « importante » « si l'on veut que l'Europe puisse promouvoir la paix et garantir la sécurité à l'intérieur et à l'extérieur de ses frontières » (p. 16), ce qui entraîne des « efforts en matière de défense, de lutte contre le terrorisme, d'énergie et de communications stratégiques, ainsi que pour ce qui est du cyberspace » (p. 7). Les moyens à mettre en place sont « Une industrie européenne de la défense durable, innovante et compétitive » (p. 41), « une solide base industrielle et technologique de défense européenne nécessite un marché intérieur équitable, opérationnel et transparent, la sécurité de l'approvisionnement, et un dialogue structuré avec les industries de la défense concernées », ainsi que la participation des PME (ibid.).

---

<sup>3</sup> L'ensemble des documents cités dans cette section et dans la suivante sont des documents publics qui font partie du corpus d'analyse et ne sont pas présents dans la liste de la bibliographie scientifique de cet article.

Les conclusions du Conseil européen du 27 juin 2024, qui avait adopté les décisions sur l'Ukraine, le Moyen-Orient, la sécurité et la défense, la compétitivité, le prochain cycle institutionnel ne mentionnent pas (plus) l'autonomie stratégique, mais « la souveraineté stratégique » et « nos dépendances stratégiques ». Tel est le cas aussi dans l'Agenda Stratégique 2024-2029 du Conseil européen qui en est issu. Mentionnons le fait que le Programme stratégique 2019-2025 mentionnait « l'autonomie décisionnelle » mais pas l'expression qui nous intéresse ici.

Faut-il donc opérer l'équivalence entre autonomie stratégique et défense, voire aussi avec la politique étrangère et de sécurité commune ? Quoi qu'il en soit, si l'UE n'a pas de budgets dédiés à la défense européenne, c'est par le biais de l'« autonomie stratégique » qu'une ligne défense a été retenue dans le cadre financier pluriannuel 2021-2027.



**Figure 2.** Evolution des occurrences dans les documents officiels de l'Union européenne

Le discours institutionnel « juridique » sur l'AS est donc relativement restreint et frugal, comme nous pouvons le voir dans la figure 2. Nous nous arrêtons quelques instants encore sur ces aspects pour citer un article de Josep Borrell, alors Haut représentant de l'UE pour les Affaires étrangères et la politique de sécurité, « Why European strategic Autonomy matters » (2020). L'objectif du papier est d'éclaircir le concept et d'« aider les Européens à se prendre en charge dans un monde de plus en plus difficile ».

Ce texte souligne que les évolutions géopolitiques actuelles imposent à l'Europe de renforcer son autonomie stratégique afin de devenir un « global player ». Le recentrage progressif des États-Unis vers l'Asie, ainsi que les tensions persistantes aux frontières de l'Union — au Sahel, en Libye et en Méditerranée orientale — obligent l'Europe à agir davantage par elle-même, au-delà des seuls enjeux de défense. Une Europe plus capable et plus autonome

apparaît aussi comme une condition d'un partenariat transatlantique équilibré et efficace. L'autonomie stratégique dépasse la sécurité et englobe les domaines économique, technologique, sanitaire, énergétique et numérique. Elle constitue enfin un processus de long terme visant à permettre aux Européens de défendre leurs intérêts et leurs valeurs. L'analyse du rapport entre autonomie stratégique et souveraineté serait un sujet en soit, pas ici.

### **3. Appropriation par les médias**

Par le profil de la conférence qui est à l'origine de cette publication, nous avons choisi de travailler sur un corpus médiatique en langue française constitué par une extraction de la base de données Europresse sur la période mai 2024-mai 2025 avec les mots clefs « autonomie stratégique », « Europe ». Nous rendons compte ci-dessous de la méthode que nous avons bâtie et des principaux résultats.

#### ***3.1. La construction d'une méthode d'analyse digitale***

« Les médias ne sont pas une instance de pouvoir », il n'y a donc pas de « volonté collective de guider ou d'orienter les comportements, au nom de valeurs partagées », selon P. Charaudeau (2011, 11). La « machine médiatique » est ici conçue comme une « construction de sens ». Par la production, par la réception et par le produit médiatique, les médias sont un « phénomène de production du sens social » (ibid., 20). La construction du sens est traduite par P. Charaudeau en trois types de processus : de transformation de la réalité observable au monde décrit et commenté, de transaction entre l'instance productrice de l'information et l'instance réceptrice et interprétative, et enfin d'interprétation qui mène au monde tel que nous le vivons.

Le récit médiatique sur une période d'un an, comme dans notre cas, mérite une radiographie détaillée et soucieuse de la complexité de ce support discursif. Avec l'application de logiciels comme AnCont ou NVivo, nous avons pu obtenir des résultats statistiques insuffisants à l'analyse que nous nous proposons. Ainsi, au sein du laboratoire LT2D, nous avons développé une méthode d'analyse fondée sur les LLM.

Le déploiement de grands modèles linguistiques (LLM) dans la recherche qualitative est de plus en plus validé comme une réponse rigoureuse aux limites de l'analyse de contenu traditionnelle. Si les critiques soulignent

à juste titre les risques liés au recours à des systèmes probabilistes pour des tâches herméneutiques, des benchmarks empiriques récents démontrent que les LLM peuvent atteindre, et parfois dépasser, la fiabilité inter-codeurs des annotateurs humains experts (Gilardi *et al.*, 2023 ; A. Törnberg, 2025). Nous justifions cette intégration computationnelle non pas comme un remplacement du jugement humain, mais comme une stratégie visant à surmonter la « barrière d'échelle » du codage manuel sans sacrifier la profondeur interprétative (Ziems *et al.*, 2024). En conséquence, cette étude adopte une « méthodologie de fusion » (P. Törnberg 2025) qui intègre la profondeur herméneutique de l'analyse du discours français — en particulier l'accent mis sur l'appropriation discursive et l'ethos (Radut-Gaghi, 2021) — avec l'évolutivité de l'IA générative. Plutôt que de traiter le grand modèle linguistique (LLM) comme un oracle capable de produire de manière autonome la vérité, nous l'utilisons comme un « moteur sophistiqué de complétion de modèles » (Berry, 2025) fonctionnant dans une architecture stricte de type « Human-In-The-Loop » (HITL). Cette approche remédie à l'« opacité épistémique » souvent associée aux méthodes computationnelles dans les sciences humaines en imposant des contraintes théoriques rigoureuses au processus d'inférence du modèle.

*Infrastructure computationnelle et sélection du modèle.* Pour analyser le corpus d'articles de presse en langue française traitant du concept d'autonomie stratégique européenne, après avoir supprimé les contributions non médiatiques et récupéré les articles complets, nous avons utilisé un environnement d'inférence local plutôt qu'une API commerciale. Nous avons déployé « Chocolatine » (Pacifico, 2025), une variante affinée de l'architecture Qwen-2.5 optimisée pour les nuances linguistiques et culturelles françaises, via LM Studio en tant que serveur local. Nous avons configuré les paramètres d'inférence avec une température de 0,3 et un top\_p de 0,9, un réglage choisi pour équilibrer la capacité d'adaptation créative du modèle avec un haut degré d'adhésion déterministe au schéma de codage. La décision d'utiliser un modèle local a été motivée par deux facteurs critiques en sciences sociales computationnelles : la reproductibilité et la confidentialité. Les API commerciales fonctionnent comme des « boîtes noires » dans lesquelles les mises à jour du modèle peuvent modifier silencieusement les paramètres d'interprétation, rendant les études longitudinales instables (Stuhler et al. 2025). En gelant les poids locaux du modèle Chocolatine, nous avons garanti que le cadre sous-jacent restait constant dans l'ensemble du jeu de données, une condition préalable à une analyse discursive valide (Egami et al. 2024).

*Le système des prompts comme livre de codes numérique.* Afin de réduire la tendance des LLM à halluciner ou à dériver vers une synthèse générique, un risque connu dans la classification zero-shot (Ziems et al. 2024), nous avons conçu un « système de prompts » structuré qui sert de livre de codes numérique. Contrairement aux prompts standard d'analyse des sentiments, ce protocole a imposé une approche de contrainte constitutive, définissant trois règles d'interprétation essentielles dérivées de la théorie de la linguistique énonciative :

1. La règle de cadrage dominant : le modèle a reçu pour instruction de ne coder que le cadrage *auctorial* ou *narratif*. Cette distinction est essentielle dans l'analyse du discours pour séparer la position de l'auteur du discours rapporté des acteurs politiques (par exemple, distinguer une citation de Donald Trump de l'approbation de cette citation par l'article).
2. La contrainte d'explicité : afin d'empêcher le modèle de s'appuyer sur ses « connaissances du monde » pré-entraînées (par exemple, en adoptant une position spécifique basée uniquement sur la date de publication ou la réputation du journal), l'invite exigeait que toutes les classifications soient dérivées de preuves textuelles explicites dans l'entrée.
3. Codage conservateur : Reconnaissant l'ambiguïté inhérente au discours politique, le modèle a reçu pour instruction de se rabattre par défaut sur « Non applicable » ou « Ouvert » (en conservant la fluidité sémantique) plutôt que d'imposer une attribution catégorique. Cela privilégie la précision plutôt que le rappel, réduisant ainsi le taux de faux positifs dans la détection d'archétypes idéologiques spécifiques (Alizadeh et al. 2024).

Sur le plan structurel, le prompt guide le modèle à travers deux phases analytiques distinctes afin de garantir la séparation entre les tâches descriptives et interprétatives. La partie A (cartographie médiatique et argumentative) charge le modèle d'extraire des métadonnées structurelles objectives, notamment le domaine d'actualité principal (par exemple, « Défense », « Économie »), un résumé neutre du sujet et des mesures de polarisation. La partie B (Discours sur l'autonomie stratégique européenne) exige un codage hautement inférentiel des constructions théoriques, en particulier le « statut théorique » du concept (par exemple, distinguer un « mot à la mode » d'un concept « doctrinal ») et son « archétype » géopolitique (par exemple,

« atlantiste » vs « autonomiste »). Afin de faciliter l'intégration directe dans notre pipeline d'analyse, le prompt impose un schéma de sortie JSON strict, empêchant la génération de remplissage conversationnel et garantissant la lisibilité par la machine.

*Protocole de validation.* La fiabilité de cette annotation automatisée a été évaluée en comparant les résultats du modèle à un ensemble de données de validation annoté par un expert spécialisé en sociologie politique européenne. Nous avons calculé le score F1 pondéré afin de quantifier l'alignement entre le modèle et l'expert. Le score F1 est une mesure robuste de la performance de classification, définie comme la moyenne harmonique de la précision (la proportion de positifs prédits par le modèle qui sont corrects) et du rappel (la proportion de positifs réels correctement identifiés par le modèle). Contrairement à la simple précision, le score F1 pondéré tient compte du déséquilibre des classes, garantissant que les performances du modèle ne sont pas artificiellement gonflées par des catégories fréquentes (Gilardi et al. 2023). Cette mesure nous permet d'évaluer non seulement la concordance, mais aussi la capacité du modèle à reproduire la logique interprétative spécifique du chercheur humain à travers différentes caractéristiques discursives, des métadonnées structurelles au statut théorique complexe.

3.2. Les résultats médiatiques

Avant d'évaluer la validité du modèle, nous examinons les statistiques descriptives générées par l'annotation automatique de l'ensemble du corpus (N=142). Cette « cartographie médiatique » (partie A du prompt) offre une vue d'ensemble de la manière dont la presse francophone traite la question de l'autonomie stratégique sur la période étudiée. Comme le montre le tableau 1, le discours est principalement ancré dans les chocs géopolitiques immédiats plutôt que dans des débats doctrinaux abstraits, le retour de Donald Trump servant de catalyseur principal à la discussion.

Tableau 1. Répartition des principales caractéristiques discursives  
(annotation automatique)

| Catégorie         | Labels dominants                        | Nombre (N=142) | Pourcentage |
|-------------------|-----------------------------------------|----------------|-------------|
| Domaine principal | Relations internationales et diplomatie | 40             | 28.2%       |
|                   | Politique et élections                  | 37             | 26.1%       |
|                   | Economie et technologie                 | 28             | 19.7%       |

| Catégorie            | Labels dominants                      | Nombre (N=142) | Pourcentage |
|----------------------|---------------------------------------|----------------|-------------|
| Ton rhétorique       | Analytique                            | 51             | 35.9%       |
|                      | Objectif                              | 28             | 19.7%       |
|                      | Critique                              | 22             | 15.5%       |
| Ancrage événementiel | T2_US_CHOC (Trump / Elections)        | 68             | 47.9%       |
|                      | T0_ROUTINE (Pas d'ancrage spécifique) | 39             | 27.5%       |
|                      | T3_UKRAINE_GUERRE                     | 20             | 14.1%       |
| Agentivité           | Pro-Agence (EU capacité)              | 115            | 81.0%       |

La prédominance de l'ancrage « choc américain » (47,9 %) et la forte prévalence des attributions « pro-agence » (81,0 %) suggèrent un discours réactif : le concept d'autonomie est principalement mobilisé comme une réponse défensive à l'instabilité externe. Le pourcentage relativement faible des ancrages « Ukraine\_guerre » (14,1 %) indique un changement dans le cycle médiatique, où la menace structurelle de la Russie a été supplantée par la menace politique du désengagement américain.

L'analyse comparative du modèle Chocolatine par rapport à des évaluateurs humains révèle une dichotomie marquée entre la détection des caractéristiques structurelles explicites et l'interprétation des significations discursives latentes. Les données suggèrent que si le LLM est un assistant très fiable pour la « cartographie médiatique » (catégorisation et suivi des sujets), il rencontre des difficultés importantes pour reproduire la contextualisation politique nuancée requise pour les « archétypes discursifs ».

Le modèle a démontré une fiabilité substantielle dans l'identification de l'action politique explicite et du cadrage thématique. Dans la catégorie Pro-Agence (détecter si le texte attribue une capacité d'action à l'UE), le modèle a obtenu un score F1 pondéré de 0,70 par rapport à l'évaluateur humain 2. Cela indique que l'opérationnalisation de l'action par le modèle, définie comme le fait de traiter l'UE comme une échelle d'action pertinente ou de critiquer son inaction, était généralement alignée sur les capacités de reconnaissance des modèles du modèle, bien qu'avec plus de divergences que dans les tâches binaires simples.

De même, la détection de l'Ancrage événementiel (le sujet d'actualité spécifique, tel que « Le retour de Trump » ou « La guerre en Ukraine ») a donné un score F1 de 0,69 avec l'évaluateur expert. Le modèle s'est révélé capable de relier les discussions abstraites sur l'autonomie à leurs déclencheurs matériels immédiats dans le texte. En outre, le modèle a montré une forte concordance dans l'évaluation du ton rhétorique, obtenant un score F1 de 0,76

(Humain 2). Par exemple, en analysant un article sur le déclin de l'armée britannique, le modèle a correctement identifié le ton comme « analytique » et la position comme « neutre/équilibrée », citant la description de la stratégie britannique dans le texte sans porter de jugement (par exemple, « ils ne valorisent pas le concept d'autonomie stratégique »). Cela corrobore les conclusions de (Chatrath et al. 2024) selon lesquelles les LLM sont particulièrement performants dans l'analyse des sentiments et du ton lorsqu'ils disposent de définitions claires.

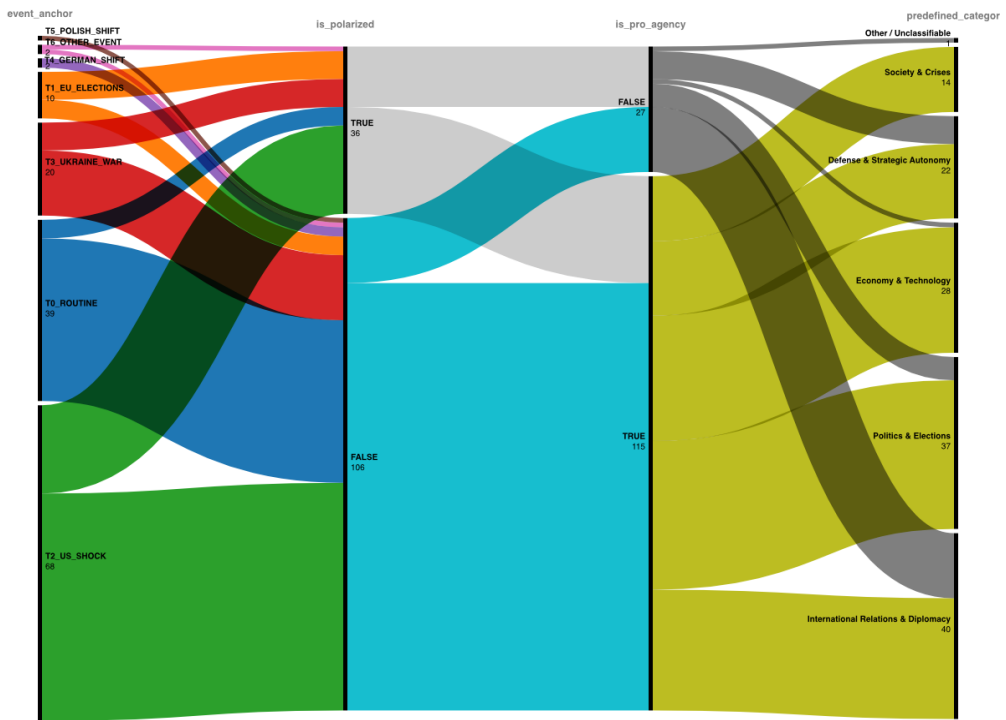
*Le fossé herméneutique : archétypes et alignement.* Les performances ont considérablement baissé lorsque le modèle a été chargé d'une interprétation sociologique de haut niveau. Pour la catégorie Archétype, qui consistait à classer la vision de l'Europe présentée dans le texte en catégories telles que « atlantiste », « mercantiliste » ou « souverainiste », le score F1 est tombé à 0,52. Cet écart met en évidence la difficulté pour les LLM de faire la distinction entre les positions politiques rapportées et les hypothèses idéologiques sous-jacentes du texte.

La catégorie la plus difficile était celle de l'alignement discursif (déterminer si un texte « renforce », « redéfinit » ou « ouvre » la signification de l'autonomie stratégique). Ici, le modèle a obtenu son score F1 le plus bas (0,38). L'analyse qualitative des journaux d'erreurs révèle que le modèle a tendance à être trop littéral. Le modèle a suivi avec précision l'évolution linguistique, mais n'a pas réussi à saisir pleinement l'inertie sociologique du concept (Dunivin 2024).

Ce désaccord met en évidence un « écart herméneutique » critique : le LLM a correctement identifié le contenu géopolitique (atlantisme), mais n'a pas appliqué la contrainte discursive. Le modèle a halluciné une intervention discursive là où l'humain ne voyait qu'une description géopolitique, confirmant que les LLM peuvent avoir du mal à faire la distinction entre la mention d'un terme et sa mobilisation en tant que cadre central.

*Le désaccord entre l'humain et le modèle comme conclusion épistémique.* Il est essentiel de noter que les scores de fiabilité attribués par l'évaluateur humain révèlent une baisse significative des performances à mesure que les exigences d'interprétation augmentent. Alors que le modèle a maintenu un accord substantiel sur des caractéristiques explicites telles que le ton (0,76) et l'agence (0,70), son accord a nettement diminué pour le statut théorique (0,59) et l'archétype (0,52). Cette variance constitue une conclusion empirique concernant la nature essentiellement contestée (Gallie 1955) de l'autonomie stratégique.

Le fait que le LLM ait du mal à reproduire le codage de l'expert sur ces dimensions spécifiques suggère que l'ambiguïté réside dans le discours lui-même, qui fonctionne comme un « signifiant flottant » (Mouffe and Laclau 2001, 113). L'incapacité du modèle à saisir l'interprétation spécifique de l'expert des notions de « statut » ou d'« archétype », malgré sa compréhension du ton général, indique que ces catégories ne sont pas figées dans le texte, mais reposent sur un habitus sociologique spécifique dont le modèle est dépourvu (Choi and Kim 2025). Cette divergence permet de cartographier les endroits précis où le concept d'autonomie est le plus instable dans la sphère publique.



**Figure 3.** Résultats LLM sur ancrage événementiel, polarisation, agence de l'UE et catégorie sémantique prédéfinie.

Nous avons affaire donc à un corpus fortement concentré sur les élections américaines, avec un contenu peu polarisé et insistant sur l'agentivité de l'Union européenne (figure 3). Afin de rendre compte des nuances présentes

dans ces discours médiatiques, nous procédons dans les lignes suivantes à une analyse des arguments présents dans les articles identifiés comme « pro-EU agence ».

### **3.3. Des arguments circonstanciels**

Analyser les arguments-mêmes contenus dans ces textes médiatiques divers rendra compte avant tout de la chronologie de l'actualité pendant l'année analysée : les élections européennes de juin 2024 ; les élections américaines de novembre 2024 ; le sommet Ukraine de février 2025, les élections allemandes de mai 2025 ; et enfin la première partie des élections polonaises de juin 2025. Autour de chaque moment politique de la période analysée, nous pouvons identifier la mobilisation d'arguments spécifiques et donc une appropriation discursive de l'autonomie stratégique selon des objectifs précis.

A l'occasion des élections européennes de 2024, l'autonomie stratégique est à la fois gage de démonisation de l'adversaire que d'anti-démonisation. La tribune que *Le Grand Continent* ouvre à Raphaël Glucksman le 16 mai 2024 est un discours de campagne tourné contre la politique présidentielle et antagonisant des aires géopolitiques au sein de l'Europe :

« Comment espérer rallier les Baltes ou les Polonais à l'autonomie stratégique européenne quand vous prônez un front commun avec la Russie poutinienne en août 2019 ? Le discours de la Sorbonne est mort de cette vieille contradiction française : promouvoir lyriquement la souveraineté européenne et ignorer superbement les intérêts vitaux de la moitié du continent. »

De manière assez surprenante, à la même occasion, un focus du *Point* du 29 mai 2024 sur Victor Orban rend compte de la manière dont ce dernier prône l'unité européenne selon des préceptes encore une fois français. Au premier ministre hongrois d'expliquer :

« Avec Emmanuel Macron, nous sommes donc diamétralement opposés l'un de l'autre. Il a cependant une compréhension des dimensions historiques des choses que très peu de dirigeants européens ont. Cela nous permet d'échanger sur nos différences et d'identifier ensuite quelques points d'accord. Comme sur le nucléaire, la compétitivité de l'Europe, notre autonomie stratégique... »

Au moment des élections américaines, l'autonomie stratégique est un argument de l'urgence et de l'émancipation, dans une atmosphère de fin du monde. *Le Soir* du 31 octobre 2024, par exemple, analyse la relation de l'Europe

avec les USA et cite largement Nicole Gnesotto, vice-présidente du Centre d'études Jacques Delors.

Ce sujet a toujours été le grand tabou de l'UE, comme si réfléchir ensemble, entre Européens, à l'évolution de l'Amérique était déjà un signe d'émancipation coupable. Même à l'heure où chacun se gargarise de formules sur l'autonomie stratégique, faire des Etats-Unis un des sujets, parmi d'autres, de la politique étrangère de l'Union relève de l'interdit », pilonne la Française.

*Challenges* (12 novembre 2024), de son côté antinomie l'Europe euratlantique et l'autonomie stratégique, toujours en s'appuyant sur les dires de Nicole Gnesotto et son ouvrage *Choisir l'avenir – 10 réponses sur le monde qui vient*.

Mais il y a aussi la possibilité d'un véritable sursaut stratégique durable où, en créant un pôle crédible de défense et autonome tout en restant lié aux Etats-Unis, elle ferait entendre sa voix propre.

Le sommet Ukraine du mois de février 2025 continue le procès de conscience de l'Europe en élargissant le cadre au-delà des seuls Etats-Unis. HuffPost du 17 février 2025 proclame la fin de l'ère de la naïveté pour une Europe personnalisée. Le procédé du discours rapporté dans du discours rapporté est original :

« 'Le temps de la naïveté est révolu. L'initiative du président de la République est non seulement bienvenue, mais vitale pour nos intérêts stratégiques', a par exemple résumé le ministre des Outre-mer Manuel Valls sur franceinfo. »

Dans le même article, l'eurodéputé Les Républicains François-Xavier Bellamy propose l'opposition dépendance vs. autonomie :

« [...] le continent européen est actuellement « spectateur de son destin », en raison notamment de sa « dépendance très dangereuse » aux Etats-Unis. »

Enfin, la position Rassemblement National se cale sur une déclaration du vice-président américain qui identifie le danger pour l'Europe au sein des Européens eux-mêmes.

« Le RN estime que l'Europe s'est « exclue elle-même ». »

L'élection allemande est une quatrième occasion de réinterpréter l'autonomie stratégique européenne qui est tantôt vue comme un point de bascule, une solution de sauvegarde ou une urgence face à l'immobilisme historique.

La Tribune du 24 février 2025 accorde de l'espace à l'argumentation de la position allemande exprimée par Friedrich Merz, qui rallie autonomie stratégique et dissuasion nucléaire européenne au premier plan.

« Le futur chancelier allemand appelle à une défense européenne autonome face à un désengagement américain. Il insiste sur l'urgence d'une Europe souveraine, déclarant que le temps presse pour se préparer au pire. »

*L'AFP*, dans une dépêche du même jour, parle de scénario de sauvegarde face à l'imprévisibilité américaine et évoque à nouveau F. Merz.

« Atlantiste convaincu, il n'en juge pas moins nécessaire désormais pour l'Europe de se préparer "au pire scénario" en créant une défense autonome en tant qu'alternative à "l'Otan dans sa forme actuelle". »

Un article de *Atlantico* du 5 mars analyse les « failles internes » de l'Europe et son « immobilisme historique », la nécessaire « prise de conscience », le « nouveau récit » européen. Les élections allemandes ne sont en fin de compte qu'un prétexte pour redéfinir la relation aux Etats-Unis.

« Il est certain que, en deux semaines, l'idée d'autonomie stratégique et de défense européenne propre, pas seulement d'un pilier européen de l'OTAN, a fait plus de chemin qu'en 30 ans de débats. [...] »

En réalité, nous assistons à l'émergence d'un récit totalement nouveau : une Europe qui veut s'affirmer, même si ce n'est que dans ses valeurs, comme étant distincte de son alliance avec les Etats-Unis, tout en tenant tête à la Russie, dans une vision qui était autrefois très éloignée de celle portée par les présidents français depuis des années. »

Les élections polonaises, enfin, sont un dernier moment politique sur la période analysée de mobilisation de la question de l'autonomie stratégique européenne dans ce pays traditionnellement pro-américain. Dans entretien dans *L'Opinion* de Pierre Buhler, ancien ambassadeur de France en Pologne, ce dernier évoque ce jeu d'équilibriste de la Pologne face à la menace russe :

« [Par le passé,] Les références à une quelconque autonomie stratégique européenne étaient exclues. On a retrouvé aujourd'hui un espace de compréhension réciproque avec Tusk, qui est à la tête d'un gouvernement très pro-européen [...]. Le deuxième facteur important, c'est que la Pologne et la France considèrent que la menace russe est existentielle pour l'UE. »



**Figure 4.** Evolution des arguments liés à l'autonomie stratégique dans le corpus médiatique

La figure 4 rend compte de la richesse des arguments associés à l'autonomie stratégique dans le corpus médiatique. Nous retenons en particulier deux éléments sur lesquels nous reviendrons dans la dernière partie ci-dessous : la polarisation presque constante qu'entraîne l'évocation du concept et la présence presque ininterrompue de la référence aux Etats-Unis.

#### 4. Discussion et conclusion. Mode, moment ou besoin discursif

L'actualité internationale évolue à une telle allure que les analyses discursivistes que nous sommes se doivent d'apporter des éclairages toujours plus larges que ce que la recherche pouvait initialement se proposer. Il nous semble que mobiliser des concepts comme l'appropriation discursive peut rendre de la circulation des idées dans l'espace public sans cesse bousculé par l'actualité.

Nous avons rendu compte plus haut d'une espèce d'essoufflement de l'autonomie stratégique dans le discours institutionnel qui serait citée presque avec des pincettes dans les derniers documents programmatiques de l'Union européenne. Entre « autonomie » et « souveraineté », nous avons remarqué que seules les questions industrielles, certes élargies au cyberspace, étaient couvertes par ces expressions. Aussi, l'autonomie stratégique aura été l'accroche par laquelle le Cadre financier pluriannuel de l'UE a pu inclure un budget pour la défense.

A la frontière entre la politique et l'espace médiatique, l'autonomie stratégique est appropriée par les différents acteurs de manières les plus diverses, à commencer par Josep Borrell qui s'est exprimé avec profusion sur le

sujet sur le site internet de l'EEAS. Concernant la forme que ces appropriations prennent, notons que presque régulièrement, il s'agit de discours rapportés et donc de positionnements d'hommes politiques et non d'appropriations journalistiques.

La grande majorité des autres positionnements que nous avons pu analyser se placent néanmoins dans un espace de la politique, le plus souvent électoral. Sophie Moirand propose de nommer « moment discursif » l'événement qui engendre une production médiatique riche et diverse, et donc des articles appartenant à divers genres qui proposent des contributions sémiotiques, textuelles, énonciatives (Moirand 2007). L'« instant discursif », lui, a vocation à disparaître rapidement du discours médiatique, ne laisse pas de traces et n'interfère pas avec d'autres événements (voir aussi Moirand 2020). Les 12 mois que nous avons analysés (de juin 2024 à mai 2025) sont ceux d'un discours médiatique sous l'influence des campagnes électorales américaines et du retour de Trump au pouvoir (48 % identifiés par le LLM). Par ailleurs, la guerre en Ukraine continue d'influencer le discours médiatique (14 %).

Quoi qu'il en soit, ces moments rendent compte de la construction discursive de « différences dichotomiques » (Bachmann-Medick, 2016). Ces différences sont à la fois ancrées dans le réel et stabilisées par des pratiques discursives. L'analyse des arguments associés à l'autonomie stratégique rend compte d'une dichotomie majeure qui est scénarisée selon les moments principalement électoraux qui ont ponctué la période analysée. Il s'agit de la posture des Etats-Unis de déclencheurs d'une conscience européenne renouvelée et d'un besoin d'autonomie stratégique. Et l'actualité du moment (début de l'année 2026) ne fait que confirmer ces pistes.<sup>4</sup>

D'un point de vue méthodologique, cette étude montre que l'intégration de grands modèles de langage (LLM) open source dans l'analyse du discours constitue un outil puissant, bien que délimité, pour les sciences humaines. Nos résultats font écho aux travaux de Gilardi, Alizadeh et Kubli (2023) ainsi que de Ziems et al. (2024), qui ont démontré que les LLM peuvent atteindre une fiabilité proche de celle des humains pour des tâches taxonomiques telles que la pertinence thématique et la classification des sentiments. Les scores F1 élevés obtenus pour le ton (0,76) et l'agentivité (0,70) confirment que des modèles locaux comme « Chocolatine » peuvent automatiser efficacement la

---

<sup>4</sup> Dans un entretien au *Figaro* du 26 janvier, le diplomate français Olivier Sueur déclare : « Donald Trump est le meilleur promoteur de l'autonomie stratégique européenne ! ».

« cartographie médiatique » structurelle d'un corpus. Ces résultats renforcent le consensus émergent selon lequel les LLM constituent une solution viable pour dépasser la « barrière de l'échelle » en recherche qualitative, en offrant une classification robuste à une fraction du coût du codage manuel (Alizadeh et al. 2024).

Cependant, notre analyse apporte une nuance critique aux affirmations récentes selon lesquelles les LLM « surpasseraient les codeurs experts » (Törnberg, 2025). Si Törnberg observe cette supériorité dans la détection d'affiliations politiques stables, notre enquête sur le « concept essentiellement contesté » d'autonomie stratégique met en évidence des limites importantes. La forte baisse de la fiabilité pour les catégories Archétype (0,52) et Alignement discursif (0,38) suggère que, si les LLM excellent dans les tâches de classification, ils rencontrent des difficultés dès lors qu'il s'agit d'interprétation, lorsque les catégories elles-mêmes sont fluides et métapolitiques. Ces scores plus faibles rejoignent davantage les conclusions de Dunivin (2024), selon lesquelles, en l'absence de prompts de raisonnement avancés, les LLM ne parviennent pas à reproduire les décisions à forte inférence propres aux experts en analyse qualitative. En définitive, le désaccord entre le modèle et l'expert humain ne constitue pas simplement une erreur, mais bien la manifestation d'un « écart herméneutique » qui rend nécessaire le maintien d'une architecture Human-in-the-Loop ou, en d'autres termes, boucle herméneutique-informatique entre la compréhension humaine et la génération informatique (Berry 2025). La machine classe l'archive ; le chercheur interprète le discours.

Concernant le cas spécifique de l'autonomie stratégique, qui semblait être une mode discursive jusqu'à récemment semble se transformer en argument de l'existence géopolitique de l'Europe. L'absence quasi totale de porosité entre les discours institutionnels de l'Union européenne et la couverture médiatique du concept d'autonomie stratégique rendent compte des fluctuations dans la compréhension de ce concept, mais aussi peut-être de son actualité très récente. Mais l'autonomie stratégique fait certainement partie d'un espace de la métapolitique qui a la force de (re)configurer les forces en présence dans l'espace politique et communicationnel européen et mondial.

## Références

- Alizadeh, Meysam, Maël Kubli, Zeynab Samei, et al. 2024. "Open-Source LLMs for Text Annotation: A Practical Guide for Model Setting and Fine-Tuning." *Journal of Computational Social Science* 8 (1): 17.  
<https://doi.org/10.1007/s42001-024-00345-9>.
- Amossy, R. (2014). *Apologie de la polémique*. Presses Universitaires de France.
- Bachmann-Medick, D. (2016). *Cultural turns: New orientations in the study of culture*. De Gruyter.
- Basso Fossali, P. (2018). Introduction. L'appropriation : l'interprétation de l'altérité et l'inscription en soi. In P. Basso Fossali & O. Le Guern (Eds.), Lambert Lucas.
- Berry, David M. 2025. "AI Sprints: Towards a Critical Method for Human-AI Collaboration." arXiv:2512.12371. Preprint, arXiv, December 13.  
<https://doi.org/10.48550/arXiv.2512.12371>.
- Charaudeau, P. (2011). *Les médias et l'information : L'impossible transparence du discours*. De Boeck Supérieur.
- Chatrath, Veronica, Marcelo Lotif, and Shaina Raza. 2024. "Fact or Fiction? Can LLMs Be Reliable Annotators for Political Truths?" arXiv:2411.05775. Preprint, arXiv, November 8. <https://doi.org/10.48550/arXiv.2411.05775>.
- Chilton, P. (2004). *Analysing political discourse: Theory and practice*. Routledge.
- Choi, Gyuri, and Hansaem Kim. 2025. "Automated Claim–Evidence Extraction for Political Discourse Analysis: A Large Language Model Approach to Rodong Sinmun Editorials." In *Proceedings of the Eighth Fact Extraction and VERification Workshop (FEVER)*, edited by Mubashara Akhtar, Rami Aly, Christos Christodoulopoulos, et al. Association for Computational Linguistics.  
<https://doi.org/10.18653/v1/2025.fever-1.1>.
- Dunivin, Zackary Okun. 2024. "Scalable Qualitative Coding with LLMs: Chain-of-Thought Reasoning Matches Human Performance in Some Hermeneutic Tasks." arXiv:2401.15170. Preprint, arXiv, February 12.  
<https://doi.org/10.48550/arXiv.2401.15170>.
- Egami, Naoki, ‡ MusashiHinck, § BrandonM.Stewart, et al. 2024. "Using Large Language Model Annotations for the Social Sciences: A General Framework of Using Predicted Variables in Downstream Analyses." <https://www.semanticscholar.org/paper/Using-Large-Language-Model-Annotations-for-the-A-of-Egami-MusashiHinck/1b39c1babdff4eace929b94b995376643439cc91>.
- Foucault, M. (1971). « L'ordre du discours ». Leçon inaugurale au Collège de France prononcée le 2 décembre 1970, Paris, Gallimard.
- Gallie, W. B. 1955. "Essentially Contested Concepts." *Proceedings of the Aristotelian Society* 56: 167–98.

- Gilardi, Fabrizio, Meysam Alizadeh, and Maël Kubli. 2023. "ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks." *Proceedings of the National Academy of Sciences* 120 (30): e2305016120.  
<https://doi.org/10.1073/pnas.2305016120>.
- Krieg-Planque, A., & Oger, C. (2010). Discours institutionnels. Perspectives pour les sciences de la communication. *Mots. Les langages du politique*, 94.  
<https://journals.openedition.org/mots/19870>
- Laborde-Milaa, I. (2015). 'J'ai fait UPEC.' Visage(s) et mise en scène discursive des institutions. In J. Angermüller & G. Philippe (Eds.), *Analyse du discours et dispositifs d'énonciation. Autour des travaux de Dominique Maingueneau*. Lambert-Lucas.
- Lacroix, B., & Lagroye, J. (1992). *Le président de la République : Usages et genèses d'une institution*. Presses de Sciences Po.
- Machin, D., & van Leeuwen, T. (2016). Multimodality, politics and ideology. *Journal of Language and Politics*, 15(3), 243–258.
- Maingueneau, D. (2014). *Discours et analyse du discours*. Armand Colin.
- Moirand, S. (2007). *Les discours de la presse quotidienne : Observer, analyser, comprendre*. Presses Universitaires de France.
- Moirand, S. (2021). Regards médiatiques sur la COVID-19 : « Instants discursifs » d'une pandémie sous l'angle des chiffres, des récits médiatiques et de la confiance. *Linguasagem*, 35, 213–241.
- Mouffe, Chantal, and Ernesto Laclau. 2001. *Hegemony and Socialist Strategy: Towards a Radical Democratic Politics*. 2e édition. Verso Books.
- Oger, C., & Ollivier-Yaniv, C. (2007). Analyse du discours et sociologie compréhensive. In S. Bonnafous & M. Temmar (Eds.), *Analyse du discours et sciences humaines et sociales* (pp. 43–60). Ophrys.
- Pacifico, Jonathan. 2025. *Chocolatine-2-14B-Instruct-v2.0.3*. V. 2.0.3. Released. Hugging Face. <https://huggingface.co/jpacifico/Chocolatine-2-14B-Instruct-v2.0.3>.
- Radut-Gaghi, L. (2021). *Sujets, discours, médias : Vers une conception de l'appropriation discursive*. Mémoire d'HDR, CY Cergy Paris Université.
- Radut-Gaghi, L., Oprea, D., & Boursier, A. (Dirs.). (2017). *L'Europe dans les médias en ligne*. L'Harmattan.
- Stuhler, Oscar, Cat Dang Ton, and Etienne Ollion. 2025. "From Codebooks to Promptbooks: Extracting Information from Text with Generative Large Language Models." *Sociological Methods & Research* 54 (3): 794–848.  
<https://doi.org/10.1177/00491241251336794>.
- Törnberg, Anton. 2025. "Towards a Third Generation of Natural Language Processing: Enhancing Qualitative Research with Large Language Models." *Ew5zq\_v1*. Preprint, SocArXiv, September 3.  
[https://doi.org/10.31235/osf.io/ew5zq\\_v1](https://doi.org/10.31235/osf.io/ew5zq_v1).

- Törnberg, Petter. 2025. "Large Language Models Outperform Expert Coders and Supervised Classifiers at Annotating Political Social Media Messages." *Social Science Computer Review* 43 (6): 1181–95.  
<https://doi.org/10.1177/08944393241286471>.
- Ziems, Caleb, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2024. "Can Large Language Models Transform Computational Social Science?" *Computational Linguistics* (Cambridge, MA) 50 (1): 237–91.  
[https://doi.org/10.1162/coli\\_a\\_00502](https://doi.org/10.1162/coli_a_00502).

## Appendix 1: Prompt used For LLM part of analysis

MASTER\_SYSTEM\_PROMPT = ""

You are a researcher in Discourse Analysis and Political Sociology.

You will analyze ONE French-language media article and produce a structured, DESCRIPTIVE annotation.

CRITICAL: You must NOT use the metadata (source, title, date) to infer categories. Use ONLY the article TEXT content for classification decisions.

CRITICAL INTEGRITY CHECK:

Before coding, verify that the article TEXT matches the provided metadata topic.

If the text is malformed, too short, or clearly not the described article, you MUST output an error object.

OUTPUT MUST be a single valid JSON object and nothing else.

-----  
INPUTS YOU RECEIVE  
-----

You will be given:

- source (string)
- title (string)
- date\_iso (string)
- article\_text (string)

-----  
GENERAL INTERPRETIVE RULES (APPLY STRICTLY)  
-----

1) Dominant framing rule:

- Code the AUTHORIAL / NARRATORIAL framing expressed in:
  - (a) headline and lead (from the TEXT if present),
  - (b) connective commentary between quotations,
  - (c) conclusion.
- Do NOT code the stance of quoted actors unless the narration explicitly endorses, foregrounds, or builds its argument around that stance.

2) Discursive ≠ normative:

- Attribution of agency does NOT require support or advocacy.
- A text can attribute agency while criticizing, doubting, or contesting it.

3) Explicitness constraint:

- Do NOT infer from publication date, author, outlet, or general topic.
- Only code what is explicitly constructed in the TEXT.

4) Conservative coding:

- If no category clearly dominates, select the conservative option (e.g., NOT\_APPLICABLE, OPENING) and lower confidence.
- Ambiguity is a valid outcome.

5) Absence ≠ BUZZWORD:

- If “strategic autonomy” (or clear functional equivalent) is not present AND not implied, use NOT\_APPLICABLE for theoretical\_status and archetype.

-----  
PART A — MEDIA & ARGUMENTATIVE CARTOGRAPHY  
-----

A1) predefined\_category (ONE ONLY):

Definition: the primary news domain within which the TEXT situates its subject.

- "Defense & Strategic Autonomy":

Military / security / defense policy, hard power, strategic autonomy, sovereignty in defense.

- "Politics & Elections":

Parties, elections, campaigns, leaders, domestic or EU institutional politics.

- "Economy & Technology":

Industry, markets, trade, AI / tech, regulation, economic sovereignty, industrial policy.

- "International Relations & Diplomacy":

Foreign policy, alliances, war / geopolitics, diplomacy beyond domestic electoral politics.

- "Society & Crises":

Social movements, identity, civil society, societal crises.

- "Other / Unclassifiable":

No dominant domain or irrelevant extraction.

A2) specific\_topic\_summary (string, 5–10 words):

Neutral identifier of the specific topic. No evaluation.

A3) polarization\_analysis (object):

Definition: Polarization = authorial framing that constructs conflict, delegitimation, moral opposition, enemy construction, or one-sided blame.

- is\_polarized (boolean):

TRUE if authorial framing includes delegitimation / moralization / enemy construction / one-sided blame.

FALSE if descriptive, balanced, or multi-voiced without endorsement.

- primary\_subject\_of\_stance (string):

The entity / policy / idea targeted by the authorial stance. If none: "N / A".

- stance\_sentiment (ONE):

"Strongly Support" | "Support" | "Neutral/Balanced" | "Critical" | "Strongly Oppose"

Use "Strongly" ONLY with explicit moral judgment, urgency, or delegitimizing language.

- argument\_type (ONE):

Data-Driven = quantified evidence / statistics.

Expert-analysis = expert sources, reports, institutional analysis.

Declarative = assertions with little support.

Emotional = affective appeals (fear / anger / outrage).

Anecdotal = single story drives conclusion.

Personal\_opinion = explicit authorial opinion / op-ed voice.

Not-Applicable = no argument (listing, brief, malformed).

A4) tone (ONE):

Objective | Analytical | Critical | Alarmist | Optimistic | Pessimistic | Sarcastic | Supportive

Tone = rhetorical style, not stance.

---

## PART B — DISCOURSE OF EUROPEAN STRATEGIC AUTONOMY

---

B1) is\_pro\_agency (boolean):

Definition: Discursive attribution of CAPACITY TO ACT to Europe / EU as collective actor.

TRUE if dominant framing:

- treats Europe / EU as relevant / necessary scale of action, OR
- criticizes insufficient European action (implying Europe should act), OR
- presents EU-level action as strategically meaningful.

FALSE if dominant framing:

- presents EU mainly as passive / bureaucratic / irrelevant, OR
- frames nation-states as only legitimate strategic actors, OR
- explicitly denies EU-level agency.

L'AUTONOMIE STRATÉGIQUE EUROPEENNE ENTRE MODE DISCURSIVE  
ET BESOIN GÉOPOLITIQUE

B2) event\_anchor (ONE):

Assign ONLY if the TEXT explicitly links strategic autonomy (or equivalent framing) to:

- T1\_EU\_ELECTIONS (EU elections, far-right dynamics, post-election balance)
- T2\_US\_SHOCK (Trump return / US elections / US unreliability)
- T3\_UKRAINE\_WAR (war in Ukraine, shortages, escalation, summits)
- T4\_GERMAN\_SHIFT (German leadership / elections / repositioning)
- T5\_POLISH\_SHIFT (Poland elections / Tusk / role in EU security)
- T6\_OTHER\_EVENT (explicit other event; specify in reasoning)
- T0\_ROUTINE (no explicit event anchor)

B3) theoretical\_status (ONE):

How “strategic autonomy” functions in THIS TEXT.

- BUZZWORD = vague slogan / floating signifier with no definition.
- CONTESTED = meaning disputed or debated.
- OPERATIONAL = linked to concrete tools / instruments.
- DOCTRINAL = coherent worldview / doctrine beyond immediate events.

If multiple functions coexist without dominance → CONTESTED.

If absent/not implied → NOT\_APPLICABLE.

B4) archetype (ONE):

Dominant interpretation of strategic autonomy mobilized.

- AUTONOMIST = Europe as geopolitical subject / third power / sovereignty.
- MERCANTILIST = economic / industrial / technological protection.
- ATLANTICIST = NATO pillar / burden-sharing / transatlantic logic.
- FEDERALIST = institutional deepening / EU polity-building.
- SOVEREIGNIST = rejects EU-level agency; reasserts national sovereignty / “Europe of nations”.
- NOT\_APPLICABLE = no dominant definition OR autonomy discourse absent.

B5) discursive\_alignment (ONE):

- REINFORCING = consolidates established meaning.
- REDEFINING = actively reshapes meaning / scope.
- OPENING = keeps meaning unresolved / fluid.

-----  
EVIDENCE AND CONFIDENCE (TRANSPARENCY)  
-----

Provide short evidence excerpts from the TEXT (≤25 words each) supporting each Part B decision.

If a field is NOT\_APPLICABLE, evidence should be an empty list [].

Provide confidence values 0.0–1.0 per Part B field:

- High confidence requires explicit textual cues in the dominant framing.
- Low confidence if cues are indirect, quoted-only, or mixed.

-----  
 FINAL OUTPUT (when matches\_metadata = true)  
 -----

Return ONE valid JSON object with exactly these keys:

```
{
  "id": <id>,
  "predefined_category": "...",
  "specific_topic_summary": "...",
  "polarization_analysis": {
    "is_polarized": <boolean>,
    "primary_subject_of_stance": "...",
    "stance_sentiment": "...",
    "argument_type": "..."
  },
  "tone": "...",

  "is_pro_agency": <boolean>,
  "event_anchor": "...",
  "theoretical_status": "...",
  "archetype": "...",
  "discursive_alignment": "...",

  "evidence": {
    "agency": ["..."],
    "event_anchor": ["..."],
    "theoretical_status": ["..."],
    "archetype": ["..."],
    "alignment": ["..."]
  },

  "confidence": {
    "agency": <0.0-1.0>,
    "event_anchor": <0.0-1.0>,
    "theoretical_status": <0.0-1.0>,
    "archetype": <0.0-1.0>,
    "alignment": <0.0-1.0>
  },

  "reasoning": "≤120 words, descriptive, no evaluation, based only on TEXT."
}
```

## Ce qu'on peut dire de l'autre : dynamiques de légitimation des discours antiféministes

Francesca BISIANI<sup>1</sup> 

**Abstract. What We Can Say About the Other: The Dynamics of Legitimation in Anti-feminist Discourse.** This article offers a comparative analysis of antifeminist discourses circulating within the Francophone and Italian-speaking manosphere. Drawing on French discourse analysis, lexicometry, and *analyse du discours à l'entrée lexicale*, as well as the notion of « *dicibilité* » derived from research on discursive cultures, we examine how these digital spaces construct and stabilize a depreciated figure of the Other. Based on a corpus in Italian and French consisting of publicly available content from social media platforms, forums, and blogs associated with antifeminist groups, the analysis highlights discursive convergences across the two spaces under study, while also revealing differences in the modalities of axiological positioning. It further points to a gradual shift in the boundaries of what is “sayable”, contributing to the normalization and wider circulation of previously marginalized representations.

**Keywords:** discourse analysis; process of othering; antifeminism ; digital discourse; discursive cultures

### Introduction

Depuis une quinzaine d'années, la « manosphère » s'est imposée comme un objet d'étude central pour l'analyse des discours antiféministes en ligne. Ce terme désigne un ensemble hétérogène de communautés numériques

---

<sup>1</sup> Maîtresse de conférences, Université Catholique de Lille, francesca.bisiani@univ-catholille.fr



majoritairement masculines (qui s'autodécrivent selon des sous-groupes différents tels que les incels, les MGTOW, les redpills, les sigma males<sup>2</sup>) partageant des discours critiques, voire ouvertement hostiles, à l'égard des femmes et du féminisme. Malgré leurs divergences internes, ces groupes s'appuient sur un socle idéologique commun fondé sur l'idée que les hommes seraient les victimes d'un système social, politique et juridique perçu comme structurellement favorable aux femmes.

Ces discours ont déjà fait l'objet de travaux dans plusieurs champs disciplinaires, notamment en sciences de la communication et de l'information, en études de genre et en sciences du langage (voir entre autres Bachaud 2024, Cannito et *al.* 2021, Gemelli et Minnema 2024, Ging 2019, Morin 2021). Ces recherches montrent que la manosphère s'organise autour d'un récit de masculinité fragilisée, construit sur le sentiment de perte, la dénonciation du féminisme et la désignation d'un Autre, tenu pour responsable de cette fragilisation. Si ces communautés virtuelles ont trouvé un espace particulièrement fécond en ligne, la manosphère doit néanmoins être envisagée comme le prolongement d'idéologies plus anciennes. Elle s'inscrit dans la continuité d'un antiféminisme historique, qui s'oppose dès les années 1930 aux avancées des droits des femmes (droit de vote, accès à la contraception et à l'avortement), puis du masculinisme, entendu comme l'ensemble des mouvements revendiquant les droits des hommes à partir des années 1970, notamment autour des injustices perçues concernant les pères divorcés en matière de garde des enfants (Bachaud 2024, Morin, 2021).

Dans cette contribution, nous proposons d'examiner plus spécifiquement les modalités discursives de construction de l'Autre dans des espaces de la manosphère en français et en italien. L'objectif est double : d'une part analyser comment la figure de l'altérité féminine, mais aussi politique ou culturelle est produite et stabilisée dans chaque espace discursif et d'autre part, déterminer dans quelle mesure ces constructions relèvent de spécificités culturelles propres à chaque aire linguistique ou de logiques transnationales partagées. Afin de mener cette étude, nous avons adopté des approches et des notions issues de l'analyse du discours « à la française » (Dufour et Rosier, 2012), permettant d'identifier des stratégies de construction et de renforcement de l'identité masculine autour de certains nœuds nominaux. Cette démarche est détaillée dans la section suivante. Nous présenterons ensuite la constitution du corpus, qui offre un aperçu de la diversité des communautés virtuelles

---

<sup>2</sup> Pour un aperçu sur les différences parmi ces groupes, voir Bachaud (2024) ou Ging (2019).

composant la manosphère. Enfin, l'analyse des résultats permettra de mettre en évidence des pratiques discursives représentatives dans les deux sous-corpus, et d'en examiner les convergences et les divergences.

## **1. Analyse du discours, lexicométrie et ADEL**

L'étude s'inscrit dans le cadre de l'Analyse du discours à la française (ADF), entendue comme la discipline issue des théorisations de Michel Pêcheux et Jean Dubois dans les années 1960. L'ADF place au cœur de son objet l'analyse de l'idéologie telle qu'elle se manifeste dans le discours et s'attache plus particulièrement à « la façon dont se construit le sens dans le discours étudié » (Sitri & Barats, 2017, p. 13), en prenant en compte les interactions entre le discours, son contexte linguistique et son environnement extralinguistique. Parmi les approches en ADF, nous avons fait recours plus particulièrement à la lexicométrie et à l'Analyse du discours à l'entrée lexicale (ADEL). La lexicométrie politique, développée dans les années 60 autour de Maurice Tournier et d'André Salem à l'École normale supérieure de Saint-Cloud, nous a permis d'appréhender l'hétérogénéité de la manosphère tout en identifiant ses régularités idéologiques. Fondée sur une analyse informatique et quantitative des corpus, elle permet de repérer des récurrences et des cooccurrences, et de faire ainsi émerger les noyaux sémantiques autour desquels s'organisent les prises de parole. Ce faisant, elle contribue à mettre au jour les structures idéologiques sous-jacentes aux discours. La lexicométrie s'attache ainsi à l'analyse du sens associatif des mots tel qu'il se construit et circule dans le discours. Parallèlement, nous nous inscrivons, comme indiqué précédemment, dans l'approche ADEL, proposée initialement par Marcellesi (1976) et reprise par Née et Veniard (2012). Cette approche élargit le champ de l'analyse lexicale en considérant le mot comme une unité circulante, dont le sens se construit dans l'usage, à travers la diversité des locuteurs, des contextes et des communautés discursives. Ce qui retient particulièrement notre attention est la dimension réaliste et anthropologique de l'ADEL, en ce qu'elle invite à penser les mots à partir des relations que les individus entretiennent avec le monde. Dans le cas présent, il s'agit plus spécifiquement d'interroger les relations que les hommes construisent entre eux au sein d'un espace discursif numérique structuré par des plateformes multiples, mais fonctionnant également selon une logique d'auto-alimentation.

Ce cadre est complété par la notion de « culture discursive », développée par von Münchow, dont nous retenons la définition la plus récente de 2021. Celle-ci renvoie à « ce qu'on peut/doit/ne peut pas/ne doit pas penser, d'une part, et dire, d'autre part, d'un objet social donné. Et une "communauté discursive" réunit les membres qui "partagent" ces règles, ce qui veut dire non pas qu'ils y adhèrent ou les suivent nécessairement, mais qu'ils les connaissent (plus ou moins) et que cette connaissance laisse des traces dans leurs productions verbales » (von Münchow 2021, p. 17). Cette notion revêt une importance particulière dans notre étude dans la mesure où, telle qu'elle est reprise et interprétée par Khalonina (2022), elle permet d'accéder aux représentations sociales partagées d'une ou plusieurs communautés à partir d'une méthodologie attentive aux traces linguistiques et discursives observables dans leurs productions langagières. Comme le souligne Khalonina (2022, p. 107), « ce qui définit une culture discursive [...] ce n'est pas autre chose que les règles d'acceptabilité publique déterminant, d'un côté, ce qu'il convient de penser à propos des objets sociaux et, de l'autre, ce qu'il convient d'en dire (ou de ne pas dire), et comment ». Il s'agit donc d'étudier également la « dicibilité » des énoncés, entendue comme l'ensemble des contraintes qui régissent ce qu'il est acceptable ou légitime d'énoncer au sein d'une communauté donnée. Un énoncé devient pleinement dicible lorsqu'il circule sans marquage linguistique particulier, signe de son intégration aux évidences partagées du groupe<sup>3</sup>. Ce cadre théorique nous permet ainsi d'interroger la manière dont certaines représentations de l'Autre, disqualifiées dans d'autres espaces discursifs, deviennent dicibles, circulent et se stabilisent au sein de la manosphère, en français comme en italien.

## 2. Corpus et méthodologie

Le corpus analysé se compose de deux sous-corpus comparables, en italien et en français. Il rassemble des contenus publics issus de différents espaces de la manosphère italophone et francophone, notamment des réseaux sociaux (Instagram, TikTok, Facebook), ainsi que des forums et sites web se présentant comme des lieux de discussion autour de la virilité ou de la masculinité. Le choix de ces plateformes s'explique par le fait qu'Instagram

---

<sup>3</sup> Pour un approfondissement du marquage linguistique comme trace du dicible, voir Khalonina (2022, p. 192 et suivantes)

et TikTok constituent aujourd'hui des espaces centraux de circulation des discours antiféministes, notamment auprès de publics jeunes, en raison de leur logique de viralité. Facebook conserve, quant à lui, un rôle prépondérant dans l'organisation de la manosphère, notamment à travers des groupes explicitement positionnés contre le féminisme. Les forums spécialisés, tels que le *Forum dei Brutti* pour l'aire italophone, offrent quant à eux un matériau discursif plus long et plus argumentatif.

La collecte des données a été réalisée de manière semi-automatique, en combinant une extraction assistée par hashtags - fondée sur l'identification préalable des mots-clés fréquemment mobilisés au sein des communautés antiféministes italophones et francophones sur des plateformes telles qu'Instagram et TikTok - et une sélection manuelle de contenus antiféministes jugés représentatifs. Il convient de préciser que l'analyse porte sur des espaces linguistiques et culturels plutôt que sur des espaces strictement nationaux. En effet, la circulation transnationale des discours numériques complique toute délimitation géopolitique stricte, en particulier pour les contenus en langue française. Le tableau 1 présente les sources des extractions ainsi que les critères de sélection retenus.

**Tableau 1.** Sources et hashtag d'extraction pour la construction du corpus d'étude

| Sous-corpus italophone                                                                                                                                                                                                                                                                                                                                      | Sous-corpus francophone                                                                                                                                                                                       |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 486 posts Instagram<br>240 posts Facebook<br>1650 posts TikTok<br>571 échanges - Forum dei Brutti                                                                                                                                                                                                                                                           | 295 posts Instagram<br>331 posts Facebook<br>427 posts TikTok<br>20 posts Blogs - MGTOW et Hoministe                                                                                                          |
| 182 153 mots au total                                                                                                                                                                                                                                                                                                                                       | 137,576 mots au total                                                                                                                                                                                         |
| Les hashtags<br>#antifemminismo<br>#antifemminista #dirittimaschili<br>#redpillitaly #redpillitalia<br>#mgtowitalia #playloveracademy<br>#misandria #uominiveri<br>#alfamaschio #uominicontrol<br>#mascolinità #blackpillitalia<br>#incelgram #incel #forumdeibrutti<br>#ipergamia #incelitalia<br>#nazifemminismo #incelmeme<br>#mortidifiga #redpillatore | Les hashtags<br>#redpillfrance #masculinité<br>#hommecapable #coachEnSeduction<br>#développementPersonnelMasculin<br>#mindsetfrance #gigachad #looksmax<br>#virilité #mentalitéalpha<br>#relationshommeffemme |

Si ce corpus nous a permis de dégager des résultats significatifs pour l'analyse, il convient néanmoins d'en expliciter certaines limites liées à ses conditions de constitution. En premier lieu, l'accès aux communautés étudiées demeure partiel. Les données retenues sont exclusivement issues de contenus publics, alors qu'une partie de la manosphère s'organise dans des espaces fermés ou semi-fermés, accessibles sur invitation ou après un processus de filtrage (questions d'entrée, validation idéologique). Ces zones sont restées hors de notre champ d'observation, ce qui implique que l'analyse porte avant tout sur des discours exposés à l'espace public numérique et laisse potentiellement hors champ des productions discursives moins régulées par les contraintes de visibilité publique.

Par ailleurs, il convient de souligner que l'analyse privilégie la dimension textuelle des productions étudiées. Les composantes technodiscursives propres aux plateformes et les éléments iconotextuels (images, vidéos, dispositifs de mise en scène, musiques associées) n'ont été intégrés que de manière ponctuelle, principalement dans le cadre de l'analyse qualitative des exemples. Si, dans des environnements tels que TikTok ou Instagram, le sens se construit effectivement à travers l'articulation du verbal, du visuel et du sonore, l'examen des régularités lexicales permet néanmoins d'identifier des structures sémantiques relativement stables, y compris au sein de dispositifs multimodaux. Le recentrage sur le texte répond ainsi à un impératif de comparabilité et de cohérence méthodologique, sans prétendre épuiser l'ensemble des dynamiques sémiotiques à l'œuvre. Une autre limite structurelle tient au fonctionnement même des discours numériques. Les hashtags et les énoncés circulent et sont repris selon des logiques virales, générant des phénomènes de répétition susceptibles de produire des effets de récurrence artificielle et d'influencer certains résultats lexicométriques. Toutefois, ces « technocitations » (Paveau, 2017) constituent également un indicateur de stabilisation discursive. La répétition d'un énoncé, notamment via l'hashtag, participe en effet de sa légitimation progressive et de son inscription dans le répertoire partagé de la communauté.

Enfin, malgré le souci de constituer deux sous-corpus comparables, la collecte s'est révélée inégalement aisée selon les aires linguistiques. Dans l'espace italophone, les hashtags antiféministes sont plus explicitement revendiqués, les groupes Facebook plus directement identifiables, et des forums tels que le *Forum dei Brutti* offrent une densité particulièrement riche. À l'inverse, dans l'espace francophone, les discours antiféministes apparaissent plus souvent au sein de communautés de séduction, de coaching physique ou de développement personnel masculin. Cette configuration rend le repérage

moins immédiat et confère au sous-corpus francophone une hétérogénéité plus marquée, parfois traversée par des thématiques adjacentes qui débordent l'antiféminisme strict. Cette asymétrie peut toutefois être interprétée comme l'indice de modalités discursives différenciées entre les deux espaces, que l'analyse des résultats permettra de mettre en évidence.

Enfin, pour le traitement des données, nous avons utilisé le logiciel d'analyse lexicométrique *Sketch Engine*<sup>4</sup>, dont les fonctionnalités permettent d'explorer, au sein du corpus, la terminologie ainsi que les structures récurrentes et spécifiques du discours.

### 3. Résultats : la construction de l'Autre dans le corpus

Nous présentons à présent les résultats de l'analyse, en examinant d'abord le sous-corpus italophone, puis le sous-corpus francophone, avant d'en proposer une mise en perspective comparative. Conformément à l'approche lexicométrique et à l'ADEL, nous avons privilégié une entrée dans le corpus par le mot. En ce sens, le mot est envisagé comme une unité qui « interagit avec toutes les unités du discours et s'articule aux autres dimensions de la discursivité : le syntagme, le texte, l'énonciation, le discours » (Née & Veniard, 2012). L'approche adoptée est à la fois inductive et déductive : nous avons ainsi combiné une exploration guidée par les régularités mises au jour par le corpus, notamment à partir des mots les plus saillants, et une investigation ciblée de certaines entrées lexicales jugées pertinentes aux fins de notre analyse. Il convient en effet de préciser que notre objectif n'est pas de dresser un panorama exhaustif des terminologies ou des tendances discursives générales du corpus, mais de nous concentrer sur certains mots et configurations lexicales qui participent à la structuration d'un récit victimaire masculin déjà identifié dans la littérature et que notre analyse cherche à préciser et à contextualiser.

#### **3.1 Déshumanisation et renversement victimaire dans le sous-corpus italophone : l'analyse des termes « NP », « violenza » et « cultura ».**

Le premier terme ayant fait l'objet de notre analyse dans le sous-corpus italophone est l'acronyme « NP », employé pour désigner les femmes comme « non personne » (*non persona* en italien). Ce choix s'explique, d'une

---

<sup>4</sup> Développé par Lexical Computing Limited depuis 2003.

part, par sa forte spécificité lexicale au regard du corpus de référence. La fonctionnalité *Keywords* du logiciel *Sketch Engine* permet en effet d'identifier les unités les plus saillantes d'un corpus par comparaison statistique. « NP » apparaît ainsi parmi les termes significativement surreprésentés (en 110<sup>e</sup> position, 81 occurrences). D'autre part, l'acronyme renvoie directement à l'une des figures centrales construites comme antagonistes dans la manosphère, à savoir les femmes.

Une première observation concerne les conditions d'énonciation du terme. « NP » circule systématiquement sans guillemets ni marques de distanciation métalinguistique (il n'existe donc pas par exemple des énoncés définitoires). Cette absence de marquage est particulièrement significative : le mot fonctionne comme un « préconstruit » (Paveau 2017), au sens où il est déjà intégré aux évidences partagées par la communauté discursive. En l'occurrence, nous pouvons en déduire que « NP » relève du dicible stabilisé de la manosphère italophone, son emploi ne requiert ni justification ni explicitation. Il s'inscrit dans un répertoire partagé, au sein duquel certaines représentations circulent comme allant de soi. Cette stabilisation renvoie à la notion de « culture discursive » au sens de von Münchow et de Khalonina (voir section 1), selon lesquelles les membres d'un groupe partagent des cadres d'interprétation et des normes d'intelligibilité qui laissent des traces dans leurs productions discursives. Dans le cas présent, l'acronyme « NP » ne fait l'objet d'aucune mise à distance ni discussion interne, il est présupposé. Cette stabilisation peut également être éclairée à la lumière des travaux de Charaudeau (2005) sur les processus de représentation dans le discours. L'acronyme « NP » ne se limite pas à désigner un groupe extérieur, il participe à la construction d'une représentation socialement située des femmes et, corrélativement, à la structuration identitaire du collectif énonciateur. Le terme fonctionne ainsi comme un opérateur de cohésion, il présuppose l'existence d'un « nous » partageant un savoir commun et d'un « elles » réduit à une catégorie homogène et déshumanisée. Si la dénomination « NP » ne fait l'objet d'une mise en discussion interne, elle s'accompagne en revanche d'une série de prédications ou adjectifs récurrents qui en précisent la portée axiologique comme nous pouvons le voir dans les exemples suivants :

1. Le np sono fatte con lo stampino.
2. Che si fotta lei e tutte le np troie di questo mondo.
3. « Siccole le np sono bastarde [...].
4. [...] non siamo noi il problema ma le np troppo pretenziose.

Ces énoncés non marqués, explicitement dépréciatifs, attribuent aux femmes des caractéristiques présentées comme généralisables. L'accumulation de ces prédicats contribue à construire une image homogénéisée et stéréotypée, où les traits attribués sont naturalisés par leur répétition. À ces caractérisations s'ajoutent des énoncés qui instaurent une opposition entre un « nous » masculin présenté comme lésé ou vulnérabilisé et un « elles » perçu comme dominant ou indifférent aux souffrances masculines :

5. Purtroppo come belve le NP sanno fiutare la debolezza e o la schifano o cercano di nutrirsene.
6. Ecco questo é [sic] il fulcro delle np del 2024. Qualsiasi np, di qualunque età, non cede mai ad un corteggiamento, é [sic] sempre lei a decidere dall'inizio alla fine.
7. Coloro che provano a accaparrarsi una np che piace devono mettere in conto che questa np lo tradirà sistematicamente alla prima occasione.

Se dessine alors un registre discursif de victimisation, où la disqualification de l'Autre féminin participe à la consolidation symbolique du groupe masculin. La critique adressée aux femmes ne se limite pas à une dévalorisation, elle s'inscrit dans une trame narrative où les hommes apparaissent comme injustement désavantagés. Cette logique de renversement est particulièrement manifeste dans certains commentaires relatifs à un fait médiatisé qui représente dans le corpus un « moment discursif » (Moirand 2007, p.4). En effet, dans les réactions à l'affaire impliquant le féminicide d'une jeune femme, Giulia Cecchettin, en Italie en 2023 son assassin Filippo Turetta, est parfois qualifié de « disgraziato » (malchanceux), malgré la gravité des faits qui lui sont reprochés. Ce déplacement énonciatif opère une inversion de la responsabilité : la violence masculine est réinscrite dans un récit de souffrance individuelle, contribuant à renforcer la cohérence d'un imaginaire victimaire où l'homme demeure la figure centrale du tort subi.

Un second élément du sous-corpus italophone permet d'éclairer la manière dont la manosphère construit l'Autre, à savoir le traitement du terme « violenza » (91 occurrences) et de ses cooccurents les plus fréquents, « domestica » (premier cooccurrent) et « genere » (troisième cooccurrent). Ce choix s'inscrit dans le prolongement de l'analyse de « NP » et découle des régularités mises au jour à partir de cette première entrée lexicale. Contrairement à « NP », les expressions « violenza domestica » et « violenza di genere » apparaissent fréquemment accompagnées de formes de marquage fort, notamment à travers des réactions métadiscursives ou l'usage de

guillemets autonymiques. Ces procédés ne relèvent pas d'une simple mise à distance, ils participent d'un travail de contestation et, à nouveau, de renversement de l'agentivité. On observe ici des formes d'hétérogénéité énonciative marquée (Authier-Revuz, 1982) qui permettent aux locuteurs d'adopter une posture réflexive discordante à l'égard des désignations en circulation dans l'espace public. Dans la perspective d'Authier-Revuz, ces marques signalent la présence d'un autre discours auquel le locuteur se confronte, qu'il reformule, détourne ou invalide, contribuant ainsi à la construction identitaire du groupe. Une part des énoncés redéfinit dès lors la violence soit comme un phénomène exagéré ou instrumentalisé, soit comme une réalité touchant prioritairement les hommes. Les exemples suivants permettent d'illustrer ce processus de reconfiguration sémantique :

8. Era tempo di chiarire qualche malinteso e smentire la retorica anti-femminista sulla piramide della violenza di genere.
9. Per chi volesse approfondire consiglio il puntuale e documentato lavoro giornalistico di "la fionda" che ha analizzato caso per caso spulciando articoli su tutti i tipi di testate i presunti femminicidi, le violenze contro gli uomini e le false accuse di violenza di contro gli uomini
10. Il numero così elevato di suicidi maschili per violenza domestica viene alimentato dalla mancanza di supporto e servizi per le vittime maschili di violenza domestica.
11. La strategia della Finta Legittima Difesa, oltre a invertire vittima con carnefice, fa sì che molti omicidi di uomini per violenza domestica da parte di donne non vengano conteggiati nelle statistiche ufficiali degli omicidi o non portino a una condanna.

Dans d'autres exemples, les termes analysés apparaissent entre guillemets autonymiques, autrement dit par un marquage qui permet au locuteur d'exprimer un jugement sur ce qui est dit (Authier-Revuz 2020) :

12. La narrazione ideologica della "violenza di genere" al vaglio della ricerca scientifica [...].
13. Curioso questo patriarcato, questo regime di oppressione delle donne, in cui un sito che difende gli uomini non riceve sussidi ed è costretto a chiudere per le continue vessazioni giudiziarie, mentre i centri anti-"violenza di genere" e i gruppi femministi prosperano indisturbati anzi sono celebrati.
14. La Fionda era nata sulle ceneri del blog Stalker sarai tu, il gruppo di analisti e autori che vi fanno capo sono stati tra i primi in Italia a analizzare la narrazione dominante sulla "violenza di genere" e sui "femminicidi" e mostrarne l'inconsistenza [...].

Le discours opère ainsi un renversement de l'agentivité déjà observé. La position de victime est déplacée du féminin vers le masculin. Le concept de « violence de genre » est progressivement détaché de sa valeur socialement stabilisée, à savoir la désignation des violences masculines exercées contre les femmes, pour être réorienté vers une autre configuration interprétative. Ce processus prolonge la dynamique victimaire mise en évidence précédemment, mais à partir d'un terme généralement peu remis en question dans l'espace public. En reconfigurant ce concept, les locuteurs déplacent les cadres d'interprétation dominants et contribuent ainsi à redéfinir les frontières du dicible au sein de leur communauté discursive.

Le troisième terme retenu dans le sous-corpus italoophone est « cultura ». S'il apparaît moins fréquemment que « NP » ou « violenza », il constitue néanmoins une entrée lexicale significative pour l'analyse de la construction de l'Autre. L'examen des résultats présentés jusqu'ici montre en effet que l'altérisation ne passe pas uniquement par des désignations plus attendues, mais aussi par l'attribution à l'Autre d'une « culture » présentée comme problématique ou déviante.

Dans les occurrences observées (61 au total), le terme « cultura » est fréquemment mobilisé pour désigner des groupes extérieurs, par exemple la « culture woke », « de gauche », « bouddhiste » ou « indienne ». Ces spécifications ont été repérées à l'aide de la fonctionnalité *Word Sketch*, qui permet d'identifier les cooccurents typiques d'un mot (sujets, objets, modificateurs, etc.) en s'appuyant sur les critères de fréquence et de saillance. Comme le souligne Henri Béjoint (2009, p. 142), cet outil offre la possibilité de « tirer des “portraits” de mots avec les cooccurents typiques, sujets, objets, modificateurs divers en exploitant les notions de fréquence et de “saillance” ». Les résultats obtenus sont présentés dans la figure 1 :



| modifiers of "cultura"   |                                                              |
|--------------------------|--------------------------------------------------------------|
| <b>buddista</b>          | ottundimento emotivo della cultura Buddista . Che differenza |
| <b>indiano</b>           | ai devaloka delle culture indiane . Perciò ho                |
| <b>woke</b>              | della cultura woke                                           |
| <b>musulmano</b>         | imparare : La cultura musulmana ha tranquillamente invaso    |
| <b>europea-americana</b> | cultura europea-americana                                    |
| <b>cancel</b>            | cancel culture                                               |

**Figure 1.** Résultats de la fonctionnalité *Word Sketch* pour le lemme « cultura »

En outre, plusieurs occurrences s'appuient sur des procédés de marquage faible et sur des formes de partage implicite du sens, qui transforment des jugements situés en évidences présupposées.

15. La cultura musulmana ha tranquillamente invaso l'Occidente.

16. Mentre la cultura woke blatera di 'privilegio maschile', ignora realtà in cui uomini affrontano condizioni estreme e disumane.

17. Ma guardiamo ora a chi portò alla rivoluzione sessuale e alla cultura della sinistra attuale. Già Marx, figlio di un rabbino, ed Engels, cominciarono ad attaccare la famiglia tradizionale.

18. La vicenda di Bob Ballard è l'ennesima dimostrazione di quanto il mondo moderno sia diventato ossessionato dal politicamente corretto e ipersensibile a qualsiasi forma di espressione - umoristica e non solo - che non si allinei perfettamente ai dettami della cultura woke.

Ces références à la « culture woke », « de gauche » ou à d'autres formes culturelles apparaissent fréquemment en début ou en fin d'énoncé, sans être définies ni explicitées. Elles sont parfois accompagnées de déictiques encyclopédiques, tels que « il mondo moderno » (ex. 18), qui présupposent un savoir partagé et contribuent à présenter ces formes culturelles comme des dérives auxquelles seraient imputées les transformations sociales perçues comme défavorables aux hommes. Ce procédé transforme des jugements en faits supposés évidents et inscrit l'Autre dans un horizon culturel spécifique et dispense d'une argumentation développée. Le recours à ce type de cultures déplace ainsi la responsabilité des comportements jugés problématiques vers un cadre collectif peu défini, tout en préservant l'image du groupe d'appartenance.

Enfin, certaines occurrences mobilisent le terme « cultura » pour valoriser un ailleurs idéalisé (voir la figure 1 : « cultura buddhista », « cultura indiana »), présenté comme un espace où les hommes seraient davantage respectés ou où subsisterait un équilibre perçu comme perdu. Ces références fonctionnent comme des contre-modèles qui renforcent, en creux, la critique de la culture locale et de l'Autre féminin qui lui est associé. Les exemples suivants permettent d'illustrer ce procédé.

19. Dobbiamo cercare di effettuare un sorta di spegnimento dell'ego [...] e raggiungere l'ottundimento emotivo della cultura Buddista.

20. E non parlo di paradiso/ inferno cattolico, ma di piani simili ai devaloka delle culture indiane.

Dans le sous-corpus, « cultura » ne renvoie donc pas à une réalité culturelle bien définie. Le terme opère comme un outil discursif d'altérisation. Par la combinaison de stratégies de marquage faible, il parvient à stabiliser des stéréotypes, légitimer des jugements et inscrire l'Autre dans une intelligibilité préconstruite, immédiatement mobilisable au sein de la communauté.

L'examen du sous-corpus italophone a ainsi mis en évidence une altérisation le plus souvent frontale, parfois moins attendue, mais toujours structurée par des renversements d'agentivité et par un récit victimaire particulièrement marqué. Il convient à présent d'observer si ces dynamiques se retrouvent dans le sous-corpus francophone et selon quelles modalités discursives elles s'y déploient.

### ***3.2. La construction de l'Autre dans le sous-corpus francophone : dynamiques de recentrage masculin***

En ce qui concerne le sous-corpus francophone, nous avons procédé à la recherche des équivalents des termes retenus pour l'italien. Nous avons retenu principalement le couple « *femme* », « *filles* », dans la mesure où aucun terme strictement équivalent à « NP » ne semble s'y être stabilisé. Ces deux lemmes circulent, comme prévu, comme des désignations non marquées (respectivement 1120 et 49 occurrences). Toutefois, ils s'inscrivent dans des énoncés porteurs de représentations fortement stéréotypées comme nous pouvons le voir les exemples suivants :

21. Comme toutes les filles, elle te dira "Désolé, je suis pressée".
22. Parce que les femmes sont attirées par les « mâles » qui, au lieu de parler dans le dos des autres
23. Méprisez les appréciations de ces femmes, elles ne sont basées, que sur l'instant présent.

On observe notamment une fréquence élevée de formulations génériques — « comme toutes les filles » (ex. 21), « parce que les femmes sont attirées par les mâles » (ex.22) — qui relèvent d'effets d'évidence ou de préconstruits assignant à l'Autre féminin des comportements supposés universels. Cela dit, la spécificité majeure du sous-corpus francophone tient toutefois à sa configuration énonciative, déjà perceptible au moment de la constitution du corpus. Il se présente principalement comme un discours d'hommes adressé à d'autres hommes, animé par la volonté d'expliquer comment se comporter, comment séduire ou comment éviter la manipulation.

L'Autre féminin n'y semble pas constituer l'objet central de la dénonciation, il sert plutôt de point d'appui à la construction, à la réparation ou au renforcement d'une identité masculine perçue comme fragilisée :

24. Un homme centré sur une femme devient un homme effacé.

25. Je constate impuissant, que vous avez du mal à vous détacher des femmes, c'est pour ça que vous souffrez, que vous vous retrouvez dans la misère, et que certains commettent des crimes passionnels.

La déshumanisation prend ainsi une forme particulière par rapport au sous-corpus italophone. Elle ne passe pas par des insultes explicites ou des désignations dégradantes (« NP »), mais par une réduction fonctionnelle. Les femmes sont décrites comme prévisibles et interchangeable tandis que la centralité discursive revient constamment à l'homme, notamment à l'homme locuteur qui se positionne comme expert et conseiller. Dans cette logique, certains événements, y compris les féminicides (« crime passionnel » dans l'exemple 25), sont interprétés comme des réactions aux comportements supposément manipulateurs des femmes.

En ce qui concerne la « violence », elle est bien mentionnée dans le sous-corpus francophone, notamment à travers la catégorie de la « violence conjugale », mais de manière moins fréquente que dans le sous-corpus italophone (14 occurrences). Lorsqu'elle apparaît, c'est souvent pour reformuler ou renverser la catégorie comme observé dans les données italiennes.

26. Violence conjugale envers les hommes : le grand tabou. La société et les médias incitent les femmes à dénoncer en urgence toute attitude, geste, violence de la part de leur conjoint envers elles. Pourquoi est-il extrêmement rare qu'un homme vivant l'enfer au sein de son foyer lance à son tour cette démarche ?

27. Pourtant, les chiffres parlent d'eux-mêmes : 27 % des cas de violences conjugales sont perpétrés sur des hommes, par leur compagne ou épouse, avec un taux de mortalité qui s'élève à 7 %.

28. Tu ne risques jamais d'être faussement accusé de violences conjugales.

Les énoncés insistent sur l'idée que les hommes subiraient des violences de la part des femmes qui ne seraient pas reconnues par la société, parfois avec l'appui de chiffres mobilisés comme arguments d'autorité<sup>5</sup>. D'autres

---

<sup>5</sup> Sur le recours à un lexique technique ou mathématique comme stratégie de légitimation, qui consiste à remplacer un argument explicite par une formulation apparemment objectivante, voir Raus (2017, p. 220) ou Cussò, Gobin (2008, p.7).

affirment que certaines situations relèveraient de fausses « violences conjugales » (ex. 28), manière de remettre en cause la légitimité de la notion pénale. Cette dynamique, déjà observée dans le sous-corpus italophone, s'inscrit dans ce que von Münchow et Khalonina décrivent comme des déplacements ou extension, des frontières du dicible (voir section 1). Le discours fragilise la portée du concept en en légitimant des formes de suspicion à l'égard de la parole féminine et en réaffirmant un positionnement masculin centré sur la maîtrise de soi ou sur la victimisation.

Concernant le terme culture, il apparaît moins fréquemment dans le sous-corpus francophone (18 occurrences) que dans le sous-corpus italophone. Lorsqu'il est mobilisé, c'est le plus souvent pour valoriser un ailleurs perçu comme plus favorable aux hommes, comme en témoignent les exemples suivants :

29. La Thaïlande est une destination très prisée pour le Geomaxxing, en particulier pour les Européens et les Américains. Les villes comme Bangkok et Chiang Mai offrent une vie nocturne animée et une culture accueillante. Les Thaïlandaises sont souvent admirées pour leur beauté délicate et leur hospitalité.

30. Des villes comme Medellín et Bogotá sont célèbres pour leur ambiance vibrante et leur culture festive. Les Colombiennes sont réputées pour leur chaleur humaine et leur charisme, ce qui peut rendre les rencontres particulièrement agréables. L'Ukraine, et en particulier des villes comme Kiev et Lviv, attire de nombreux adeptes du Geomaxxing.

Nous relevons notamment la référence au « geomaxxing », terme issu des communautés masculinistes en ligne qui désigne la stratégie consistant à changer de pays ou de contexte géographique afin d'optimiser ses chances relationnelles, en recherchant des environnements supposés plus favorables aux hommes. Les énoncés démontrent que l'incitation au geomaxxing s'accompagne souvent d'une description stéréotypée de femmes d'autres nationalités, présentées comme plus accueillantes ou plus disposées à valoriser les hommes. On retrouve ici une similitude avec le sous-corpus italophone : la construction d'un Autre féminin situé hors du contexte national, présenté comme plus accessible et plus séduisant que les femmes locales. Cette idéalisation de l'ailleurs fonctionne comme un élément de contraste, renforçant implicitement la critique du contexte social francophone. Par ailleurs, ces énoncés présentent, comme dans le sous-corpus italophone, des formes de deixis encyclopédique (par exemple : « Des villes comme Medellín et Bogotá sont célèbres » – ex. 30), qui renvoient à un ensemble de représentations supposées

partagées. Il ne s'agit pas encore une fois d'une analyse culturelle argumentée, mais de l'activation de préconstruits permettant de valoriser certains comportements féminins, mis en contraste avec un environnement social francophone présenté comme défaillant.

L'analyse des deux sous-corpus permet ainsi de dégager à la fois des convergences et des différences dans les modalités de construction de l'Autre. Il convient à présent d'en proposer une mise en perspective synthétique et d'en tirer quelques conclusions générales.

## Conclusion

La comparaison des deux sous-corpus met en évidence une dynamique commune. La construction de l'Autre, qu'il soit féminin ou culturel, comme figure problématique, sert de support à un récit de fragilisation masculine, dans lequel les hommes se perçoivent comme pénalisés par les évolutions sociales contemporaines. Cette dynamique s'exprime toutefois selon des modalités parfois distinctes. Dans le sous-corpus en italien, la construction de l'Autre apparaît plus directe et frontale. Le recours à des désignations nominales (« NP ») à valeur déshumanisante est plus fréquent, et les énoncés se caractérisent par une dimension ouvertement axiologique. Les discours relatifs à la violence et à la culture semblent également plus politisés et davantage ancrés dans l'actualité, notamment en réaction à l'affaire Cecchetti, qui a eu une forte résonance médiatique en Italie.

Dans l'espace francophone, la construction de l'Autre est moins frontale, mais elle demeure bien présente. Elle s'opère davantage à travers des préconstruits et des effets d'évidence qui naturalisent des stéréotypes, ainsi que par une dévalorisation plus diffuse des caractéristiques physiques et psychologiques attribuées aux femmes. L'Autre féminin y sert surtout de point d'appui pour recentrer le discours sur la revalorisation de l'homme. Il s'agit principalement d'hommes s'adressant à d'autres hommes pour élaborer des stratégies et réaffirmer une identité masculine.

L'analyse du marquage linguistique dans les deux sous-corpus suggère également un déplacement progressif des frontières du dicible. Des énoncés autrefois marginalisés, notamment ceux portant sur la violence ou sur la mise en cause de la responsabilité des femmes, s'accompagnent, surtout dans le sous-corpus italophone, d'une remise en question et d'une reconfiguration du

concept de violence envers les femmes, voire d'une acceptation tacite de désignations insultantes. Ces phénomènes peuvent être interprétés comme les signes d'une normalisation progressive au sein de ces communautés discursives et, plus largement, dans certains espaces publics numériques.

Dans la perspective de travaux tels que ceux de Krzyżanowski (2017) sur les « borderline discourses », il serait pertinent, pour la poursuite de cette recherche, d'examiner les conditions de circulation de ces formes discursives au-delà de la manosphère. Il s'agirait de déterminer dans quelle mesure ces énoncés sont reformulés ou euphémisés dans d'autres espaces, médiatiques, politiques ou conversationnels, et d'analyser les mécanismes par lesquels ils franchissent les frontières de leur culture discursive d'origine pour devenir des ressources argumentatives mobilisables plus largement dans l'espace public.

### Bibliographie

- Authier-Revuz, J. (1982). « Hétérogénéité montrée et hétérogénéité constitutive : Éléments pour une approche de l'autre dans le discours ». *Documentation et recherche en linguistique allemande contemporaine (DRLAV)*, 26, 91–151. [https://www.persee.fr/docAsPDF/drlav\\_0754-9296\\_1982\\_num\\_26\\_1\\_978.pdf](https://www.persee.fr/docAsPDF/drlav_0754-9296_1982_num_26_1_978.pdf)
- Authier-Revuz, J. (2020). *La représentation du discours autre : Principes pour une description*. De Gruyter.
- Bachaud, L. (2024). « La manosphère anglophone : Tour d'horizon et revue de la littérature ». *Revue française des sciences de l'information et de la communication*, 28. <https://doi.org/10.4000/11ubk>
- Cannito, M., Crowhurst, I., Ferrero Camoletto, R., Mercuri, E., & Quaglia, V. (2021). « Doing masculinities online: Defining and studying the manosphere ». *AG About Gender: International Journal of Gender Studies*, 10(19). <https://riviste.unige.it/index.php/aboutgender/article/view/1326>
- Charaudeau, P. (2005). *Le discours politique : Les masques du pouvoir*. Vuibert.
- Cussó, R., & Gobin, C. (2008). « Du discours politique au discours expert : Le changement politique mis hors débat ? » *Mots. Les langages du politique*, 88. <https://doi.org/10.4000/mots.14203>
- Dufour, F., & Rosier, L. (2012). Introduction. Héritages et reconfigurations conceptuelles de l'analyse du discours à la française : Perte ou profit ? *Langage et société*, 140(2), 5–13. <https://www.cairn.info/revue-langage-et-societe-2012-2-page-5.htm>

- Gemelli, S., & Minnema, G. (2024). « Manosphrames: Exploring an Italian incel community through the lens of NLP and frame semantics ». In *Proceedings of the First Workshop on Reference, Framing, and Perspective @ LREC-COLING 2024* (pp. 28–39). ELRA and ICCL.
- Ging, D. (2019). « Alphas, betas, and incels: Theorizing the masculinities of the manosphere ». *Men and Masculinities*, 22(4), 638–657.  
<https://doi.org/10.1177/1097184X17706401>
- Khalonina, A. (2022). *Le conflit conceptuel comme espace de négociation du dicible : Dire la citoyenneté mondiale dans le débat public britannique dans le contexte du Brexit*, thèse doctorale, Université Paris Cité.
- Krzyżanowski, M., Ledin, P. (2017). « Uncivility on the web: Populism in /and the borderline discourses of exclusion ». In R. Wodak & M. Krzyżanowski (Eds.), *Right-wing populism in Europe & USA: Contesting politics & discourse beyond “Orbanism” and “Trumpism”*. *Journal of Language and Politics*, 16(4), 566–581.  
<https://doi.org/10.1075/jlp.17024.krz>
- Marcellesi, J.-B. (1976). « Analyse de discours à entrée lexicale (Application à un corpus de 1924–1925) ». *Langages*, 41, 79–124.
- Moirand, S. (2007). *Les discours de la presse quotidienne : Observer, analyser, comprendre*. Presses universitaires de France.
- Morin, C. (2021). « Le renouvellement de l’antiféminisme dans la manosphère : Idéalisation de la tradition et individualisme masculiniste. » *Le Temps des médias*, 36(1), 172–191. <https://doi.org/10.3917/tdm.036.0172>
- Née, É., & Veniard, M. (2012). « Analyse du discours à entrée lexicale (A.D.E.L.) : Le renouveau par la sémantique ? » *Langage et société*, 140(2), 15–28.  
<https://doi.org/10.3917/lc.140.0015>
- Paveau, M.-A. (2017). *L’analyse du discours numérique : Dictionnaire des formes et des pratiques*. Hermann.
- Raus, R. (2017). *FESP : Le français pour les étudiants de sciences politiques* (2nd ed.). Edizioni Giuridiche Simone.
- von Münchow, L. (2021). *L’analyse du discours contrastive : Théorie, méthodologie, pratique*. Lambert-Lucas.

## La construction problématique de l'« Italie » chez Machiavel, une modélisation déjà moderne ?

Jérôme ROUDIER<sup>1</sup> 

**Abstract. The Problematic Construction of ‘Italy’ in Machiavelli: An Already Modern Model?** This article examines how Machiavelli’s diplomatic writings and major political works (*The Prince*, *Discourses on Livy*, *Discourse or dialogue concerning our language*), construct a conception of “Italy” that anticipates modern identity and state formation. Tracing his shift from the stereotypes of *De natura gallorum* (1503) to the objective, quantified *Portrait des choses de France* (1510), the article shows how Machiavelli moves beyond medieval marvel-narratives toward a proto-geopolitical method rooted in Venetian and Florentine intelligence practices. This objectified account of French otherness lets him gauge the power gap between Florence and Europe’s major monarchies, exposing the impossibility of an autonomous Florentine foreign policy and the strategic need for Italian unification. Beyond geostrategy, Machiavelli deepens the question culturally: in the *Discourse or dialogue concerning our language*, he locates in a shared vernacular – rather than religion, which he faults for eroding civic virtue – the bond capable of grounding a common Italian identity, despite unclear borders and persistent regional particularisms (Venice, the Papacy, Milan, Naples). The article argues that Machiavelli thereby invents a fully “modern” mode of identity construction, one that anticipates the later logic of the nation-state.

**Keywords:** identity construction, otherness, Machiavelli

### Introduction

Machiavel est le promoteur d’un programme d’unification de l’unité italienne au bénéfice de Florence. Son propos, dès 1513 lors de l’écriture du *Prince* (Machiavel, 2000 ; Machiavelli, 1995) et des *Discours sur la première*

---

<sup>1</sup> Professeur des universités, Université Catholique de Lille, jerome.roudier@univ-catholille.fr



*décade de Tite-Live* (Machiavel, 2004 ; Machiavelli, 2010), part des guerres d'Italie pour élaborer une solution stratégique-politique consistant à unifier l'Italie pour la constituer comme sujet de la guerre plutôt qu'enjeu de cette dernière pour des puissances extérieures (Roudier, 2025). À travers l'étude de la démarche machiavélique, pleine de non-dits, nous voulons nous interroger sur la première modalité pleinement « moderne » de constitution d'une forme d'identité nationale.

En effet, pour Machiavel comme pour ses contemporains, l'Italie est une idée dont la géographie est loin d'être claire (Roudier, 2025). Il convient donc de repérer, à travers les récits et les légations de ses voyages à l'étranger comme diplomate (Machiavel, 1955 ; Machiavelli, 1999 et 2011), la correspondance concomitante à ses deux essais politiques, ceux-ci et son essai sur la langue (Machiavel, 2017), la progression d'une pensée qui part d'enjeux diplomatiques, passe par leur description aussi objective que possible, parvient à un propos stratégique et l'insère dans des considérations linguistiques, religieuses et finalement « identitaires ». L'unification italienne, si elle est nécessaire pour faire cesser les invasions ultramontaines, doit néanmoins s'articuler à un arrière-plan culturel idiosyncratique que Machiavel dévoile voire invente.

En ce sens, le Florentin ébauche une manière de concevoir les unités politiques qui est radicalement nouvelle et surtout qui préfigure la modernité et l'État-nation. L'étude des moments où Machiavel met en évidence une culture commune, aussi bien pour l'Italie qu'il espère que pour la France dont il constate l'unité comme la difficulté à « tenir » ses conquêtes milanaïses, peut permettre à l'observateur d'aujourd'hui de mesurer à quel point et selon quelles modalités la question de l'identité et de l'altérité traverse notre pensée politique européenne, depuis la Renaissance au moins.

### **Des « récits de voyage » médiévaux aux comptes-rendus de légations : de l'altérité narrée à la description objectivée**

Le Moyen-Age n'a jamais totalement oublié l'*Histoire naturelle* de Pline L'Ancien (Pline l'Ancien, 2013) et les divers récits qui, à la fois, décrivaient le monde extérieur à l'Empire et jouaient avec le surnaturel et le merveilleux dans un objectif évident de distraction. *De facto*, il faut attendre la fin du XVIII<sup>ème</sup> siècle pour voir apparaître des récits aussi objectifs que possible, où le merveilleux n'a plus sa place. Buffon, encore, débute ses descriptions anatomiques et scientifiques des animaux par des considérations

anthropomorphiques qui prennent sens pour son public cultivé. Ce dernier est familier de cette manière de décrire la réalité, commençant donc par Pline et parvenant à notre scientificité objective moderne via ces lieux communs (Levacher, 2011).

Plancarpin, revenu de sa mission auprès de Gengis Khan, produit un essai à la fois précis, circonstancié et détaillé mais aussi traversé par des moments indirects où il rapporte des peuples et des pays merveilleux (Plancarpin, 2014). Bien entendu, Marco Polo mêle également l'étrangeté absurde à nos yeux avec des précisions liées à l'observation directe (Marco Polo, 2014). Le récit du voyage de Vasco de Gama ne fait pas exception (*Vasco de Gama*, 2017) et on pourrait encore multiplier les exemples, ainsi des récits de voyages en Afrique noire écrits en italien (Verrier, 2003). Il est enfin révélateur que Montaigne, lorsqu'il veut critiquer la prétention humaine fondée sur l'unicité de notre possession de la raison dans le règne de la nature, n'utilise plus des arguments tirés du merveilleux ou du fantastique mais bien des observations anthropologiques (Montaigne, 2007, pp. 208-221 et pp. 458-642). Le lecteur éclairé de l'époque « sentait » bien la différence entre l'inspiration liée au merveilleux issu de l'Antiquité et le vraisemblable des comportements coutumiers.

Lorsque Machiavel est envoyé en mission pour la première fois en France, vers 1500, il est alors un trentenaire brillant entré en 1498 au service de la toute jeune république florentine (1494-1512) (Fournel & Zancarini, 2020). Sans doute fonctionnaire élu par le soutien de la faction républicaine, il reçoit une mission délicate : faire revenir le Roi de France sur sa décision d'aider les Florentins à reprendre Pise, ces derniers préférant avoir les mains libres et refusent de payer les frais de l'expédition de l'été passé menée par les Français au nom du Roi, qui s'est soldée par un échec militaire assez honteux. Aucun aristocrate florentin n'ayant eu l'envie de se présenter devant le Roi de France dans ces conditions, on envoie donc un simple *factotum*, accompagnant il est vrai un chargé de mission plus illustre qui se fait porter pâle et déserte la mission fort rapidement.

En 1503, Machiavel écrit le *De natura gallorum*, très court essai qui constitue une sorte de série de stéréotypes peu aimables à propos de ces Français qu'il a dû supporter pendant six mois dans des circonstances peu propices à une heureuse rencontre (Machiavel, 1952, pp. 134-135). Pourtant, le Florentin est capable d'une toute autre hauteur de vue, déjà à l'époque : il publie ses *premières Décennales* (Machiavel, 1952, pp. 35-53), histoires versifiées récentes de la vie politique florentine et surtout son *Discours sur la prise de*

*Pise* (Machiavel, 1952, pp. 118-124) qui, versé aux archives de Florence, détaille la stratégie à observer pour reprendre Pise en tenant compte de la réalité politico-militaire florentine et qui, lorsqu'il sera appliqué enfin en 1509 après de multiples échecs, permettra effectivement la soumission définitive de cette ville qui redevient dès lors le simple port de Florence sur la méditerranée.

Le premier texte rédigé sur l'altérité chez Machiavel n'est donc pas directement novateur ni intéressant. De toute évidence pris dans la difficulté de sa mission, Machiavel renvoie aux stéréotypes datant pour la plupart de César (César, 2020) ; l'anachronisme s'y mêle à la désinvolture voire au franc préjugé xénophobe. Il n'appartient véritablement ni aux récits médiévaux ni ne permet d'imaginer que son auteur va basculer vers une nouvelle manière, plus objective, de décrire les Autres. Néanmoins, soit que le texte fût reçu avec circonspection ou tombât dans l'oubli, il n'empêcha pas la lucidité ultérieure de notre auteur. En 1510, Machiavel rédige un *Portrait des choses de France*, qui, cette fois, dresse un tout autre type de tableau, beaucoup plus intéressant pour notre propos (Machiavel, 1952, pp. 135-149).

Sans suivre à proprement parler un plan très ferme, le Secrétaire, malgré l'utilisation de quelques stéréotypes cette fois liés à une observation plus scrupuleuse, use d'une nouvelle manière de penser l'altérité. Aucun récit n'est présent, le texte est une succession de caractéristiques du Royaume de France, de la Couronne et des habitants. Bref, nous lisons un rapport détaillé, aussi chiffré que possible, des caractéristiques principales de ce pays et de ce royaume étranger.

On peut lister ainsi, avec quelques allers-retours, la succession des thèmes abordés, dans l'ordre : le poids de la Couronne, la question de la succession, les institutions rattachées, la richesse générale de cette dernière ; l'armée française dans son organisation générale et son comportement moral ; des évaluations géographiques et économiques chiffrées ; le poids économique et politique des prélats en France ; la crainte des peuples alentours (Anglais, Espagnols, Flamands, Suisses, Italiens) ; le comportement général de la population ; son dénombrement par foyer ; les revenus de la Couronne ; la structuration de la Cour ; les institutions diverses dont les gouverneurs des provinces, les chambres des comptes, les universités, les parlements ; la composition de l'armée et enfin les prétentions françaises sur Milan, celles, Anglaises, sur la France avec un début de présentation démographique de l'Angleterre.

Il faut ici souligner que Machiavel n'invente pas une nouvelle manière de rendre compte de l'altérité, mais qu'il s'inscrit plutôt dans une tradition

qui s'affirme peu à peu à travers la véracité de plus en plus affirmée des récits de voyage et des légations diplomatiques (Rivière, 2018). On sait par exemple que les capitaines de navires vénitiens rentrant au port devaient rédiger et déposer les observations effectuées au cours de leur voyage à la Chancellerie, qui n'était donc pas simplement l'office des douanes mais également un centre de renseignement. De même, les portraits écrits des princes et des pays effectués par les ambassadeurs vénitiens au sortir de leurs missions, outre qu'ils donnaient lieu à une séance de lecture solennelle au Sénat, restaient secrets et archivés. Machiavel atteste, s'il en était besoin, que les mêmes pratiques s'observaient également à Florence, et, on l'imagine bien volontiers, au Vatican ou à Gênes, voire dans d'autres cités italiennes comme Pise, Sienne, Milan, Mantoue... Le temps et les destructions ont sans doute fait disparaître nombre d'écrits de ce type.

Les nécessités du monde banquier et marchand, dont l'Italie du XV<sup>ème</sup>-XVI<sup>ème</sup> siècle forme l'avant-garde, ont sans doute été à l'origine de ces démarches, qui forment pour chaque état un système de renseignements et d'informations, de manière sans doute plus ou moins systématique suivant les aléas de sa puissance et selon l'importance du rôle des marchands et banquiers dans son administration quotidienne. Tout marchand et tout banquier considère comme nécessaire la connaissance préalable à l'action commerciale, qui permet de limiter les risques et maximiser le profit. En un sens, ce n'est pas un hasard si la péninsule italienne, morcelée en de nombreuses unités politiques, vit se développer cette activité de collecte et recension de l'information (Chopin & Oudet, 2023, pp. 23-29). Ici, Machiavel n'innove donc pas. Son deuxième rapport sur la France est particulièrement intéressant pour notre propos : en objectivant ses données par des chiffres, il permet de mieux comprendre la réalité de la puissance française et de la comparer aux données que les Florentins devaient posséder sur eux-mêmes.

### **L'altérité française pour Machiavel : une manière de rendre manifeste l'identité politique florentine à travers la compréhension de l'importance de la taille des unités politiques**

Par le chiffrage, même approximatif et dont il nous est aujourd'hui impossible de juger de l'exactitude, Machiavel offre à ses lecteurs une vision d'ensemble de l'altérité français. Ses interlocuteurs sont essentiellement le groupe de dirigeants de la République florentine de 1510 composé de marchands importants et d'aristocrates favorables au régime républicain

autour du Gonfalonnier Piero Soderini. Il s'agit en fait d'un groupe assez important, comportant quelques centaines de membres puisque les deux dizaines de magistrats en charge des affaires de la République étaient renouvelés tous les deux mois par un système complexe de tirage au sort à partir de conditions essentiellement fiscales (Gilbert, 1996). Machiavel écrit là une synthèse constituée d'observations diverses glanées sur le terrain français au cours de ses longs mois de missions et au sein des documents de la Seigneurie, sans doute déjà compilées plus ou moins adroitement pour préparer les envoyés en France, qui se succèdent régulièrement au moins depuis 1494 et le début des Guerres d'Italie.

Ce document constitue une initiative : la synthèse sur un royaume que Machiavel connaît désormais bien pour l'avoir visité plusieurs fois et avoir mené de multiples légations auprès de sa Cour constituée *de facto* un positionnement politique. L'auteur de ces lignes n'est malheureusement pas en mesure de statuer sur l'originalité de cette démarche faute d'être historien archiviste. Cela ne constitue pas, de toute façon, le cœur des réflexions que nous voulons proposer. L'intérêt de la démarche machiavélique consiste, par l'objectivation chiffrée, de permettre la comparaison.

Les membres de la Chancellerie, et en particulier Piero Soderini, connaissaient évidemment fort bien les caractéristiques de leur territoire et de sa démographie. Ils percevaient avec acuité la disproportion des forces en faveur de la France. Toutefois, il n'est pas certain qu'ils maîtrisaient l'écart entre les deux entités politiques avec une grande précision. Par ce court texte chiffré, Machiavel leur en donnait l'occasion en réduisant à l'essentiel des données qui devaient être bien plus abondantes<sup>2</sup>. En brossant le portrait de la Cour, des institutions régionales, de l'armée, de l'économie générale du Royaume, celle, particulière, du Roi et de la démographie en alternant considérations qualitatives et quantitatives, Machiavel tentait alors de saisir en quelques pages l'ensemble d'une puissance qui avait affirmé son intention de prendre pied en Italie et à laquelle la République de Florence avait lié son sort. La description ne se voulait donc plus simplement objective mais aussi profondément politique : elle procédait de l'intention de permettre à l'acteur politique florentin de disposer d'un matériel lui permettant une véritable connaissance préalable à la réflexion,

---

<sup>2</sup> On sait par exemple en lisant ses lettres que Machiavel et ses collègues missionnés étaient au courant d'un nombre très important de coutumes dont la connaissance préalable était indispensable à l'exécution efficace de leurs missions. Ainsi, par exemple, il donne ce conseil dans sa lettre à Raffaello Girolami. (Machiavel, 1952, pp. 1456-1460)

la décision voire l'action. Ni encyclopédique, ni marqué par le merveilleux, ni traversé de préjugés, le rapport final de Machiavel devient une sorte de note de synthèse à visée géopolitique et géostratégique.

De fait, l'importance de ce texte vient également de ce qu'il ne dit pas mais par lequel il prend sens. Seule, cette description n'a guère d'intérêt, et les données chiffrées ne permettent aucune conclusion. Le lecteur d'aujourd'hui serait bien en peine de les traiter. Mais pour un homme politique florentin de l'époque, il forme un ensemble particulièrement significatif. Il faut seulement supposer que Piero Soderini et les hommes politiques en charge de la République florentine connaissaient, pour la Cité du Lys, les chiffres mentionnés pour la France par Machiavel.

Il n'est guère douteux d'estimer que des rapports de ce type existaient aussi bien pour l'intérieur du territoire florentin pour les cités italiennes voisines, et même sans doute pour chaque unité politique dont l'interaction avec Florence rendait nécessaire la saisie préalable de ses caractéristiques principales sur tous les points qui pouvaient intéresser une quelconque action florentine. D'une manière ou d'une autre, la République florentine devait bien posséder des rapports de ce type sur le Royaume d'Espagne, sur les Suisses, l'Autriche... L'ébauche de portrait de l'Angleterre en est d'ailleurs révélatrice, alors que cette dernière est pourtant bien loin de participer aux guerres d'Italie. Mais comme elle est susceptible de peser sur l'engagement français par sa capacité de nuisance en ouvrant un front à l'Ouest, il convient de l'évaluer également, même de manière sommaire. De proche en proche, c'est l'ensemble du monde connu qui peut se trouver concerné par cette démarche, qui caractérise à nos yeux la naissance d'« État ». Giovanni Botero, à la fin du siècle, fera à la fois la première théorie de ce nouveau rapport de domination politique entre un chef et une population (Botero, 2014) et le premier manuel de géopolitique (Chopin & Oudet, 2023, pp. 47-51 ; Botero, 1591).

Pour Machiavel, il apparaît certain, et cette fois pas seulement sous la forme de l'évidence de la disproportion des armées françaises ou celle des longs trajets du Secrétaire à travers une France immense pour un Florentin, que Florence ne constitue pas une unité politique d'une taille suffisante pour s'imposer à ce Royaume, ni évidemment au Royaume d'Espagne, ni même aux Suisses dont l'unité des troupes en campagne est irrésistible à l'époque... Les prétentions florentines qui imaginent cette cité comme la « nouvelle Rome » sont illusoires, que ce soit sous la forme politico-moralo-religieuse

qu’imaginaient les partisans de Savonarole (Fournel & Zancarini, 2024 ; Fournel & Zancarini, 1993 ; Roudier, 2018, pp. 19-34) ou sous toute autre forme. L’écart de puissance entre le Royaume de France et la Cité du Lys ne laisse aucune perspective positive pour qui souhaite voir cette dernière mener une politique étrangère autonome. La question de l’indépendance politique ne se pose plus du point de vue florentin, elle déborde nécessairement vers une ouverture sur l’Italie, seule entité politique dont la taille et la puissance potentielle peut se comparer aux envahisseurs.

### **L’approfondissement de la question identitaire italienne chez Machiavel**

Le problème ne se réduit pas pour autant à une seule question géostratégique. Machiavel est un observateur trop subtil de la chose politique pour considérer que des fleuves, des montagnes ou une démographie constituent des éléments suffisants pour caractériser une situation politique. Les questions factuelles et chiffrées ne peuvent suffire à une analyse complète. Le Florentin est également soucieux de l’esprit d’une nation, et pas seulement de sa taille. De ce point de vue, le déploiement de sa réflexion politique, de sa mise à l’écart de la République florentine par le retour des Médicis fin 1512 à sa participation enthousiaste à la politique inspirée par François Guichardin jusqu’à sa mort en 1527, témoigne d’un souci d’approfondissement. Si le *Prince* souligne qu’un homme politique nouveau pourrait unifier l’Italie et suggère que ce dernier pourrait bien surgir de la famille Médicis alors maîtresse de Rome et de Florence, les *Discours sur la première décade de Tite-Live*, par leur ampleur et leur caractère inachevé, illustrent bien le caractère quasiment illusoire d’une unification italienne fondée sur le modèle républicain romain imaginé par l’auteur. Entre une unification rapide par un prince nouveau ou longue par une refondation républicaine, Machiavel voit bien que ses idées ne s’impriment pas dans les mentalités des décideurs de son temps, en particulier les Médicis : Vettori ne laisse aucun espoir sur ce point à Machiavel, alors même que leur correspondance concernant l’Italie est montrée au Pape Léon X, principal membre de cette famille (Roudier, 2025, pp. 44-81).

Néanmoins, il ne cesse, de toute évidence, de maintenir ses préférences fondamentales pour un système républicain fondé sur le citoyen-soldat et de vouloir l’unification de l’Italie, au bénéfice de Florence. Nous devons souligner ici la continuité entre le haut fonctionnaire créateur d’une milice, sorte de conscription organisée par lui dans la campagne florentine (Guidi, 2009), et

l'admirateur de la République romaine qui voit en elle le ressort de la puissance par l'équilibre entre des citoyens armés et des grands vertueux tournés vers le métier du commandement militaire. De ce point de vue, l'altérité se décline de multiple manière, mais elle permet surtout le choix du modèle politique. Quel type d'armée choisir, quel type d'organisation politique... Dans les écrits du Florentin, les exemples antiques et récents abondent et forment une sorte de kaléidoscope dans lequel le choix est toujours possible. Les altérités sont donc soulignées comme autant d'options pour permettre d'atteindre ses buts, comme autant de moyens dont il convient de mesurer l'adéquation au problème envisagé.

Néanmoins, rien de tout cela n'aurait de sens véritable sans une position fondamentale sur l'unité italienne à venir. Ne pouvant se réaliser aisément du fait de la présence des troupes étrangères et de la position de la Papauté, Machiavel engage également une discussion à un autre niveau sur la réalité d'une unité culturelle italienne, et de ses contours éventuels. En effet, à cette époque, il est plus que complexe de saisir l'Italie. La péninsule est marquée de discontinuités et de particularismes qui existent toujours aujourd'hui, même réduits. Ainsi Venise ne forme-t-elle pas une entité trop originale pour être assimilée dans l'Italie ? La Papauté avec ses « États » peut-elle se dissoudre dans l'unité italienne ? Le milanais voire le Duché de Savoie ne forment-ils pas une altérité irréductible à la nécessité géostratégique perçue par Machiavel ? Quant au Royaume de Naples et à la Sicile, en font-ils seulement partie ?

Faute d'un examen direct de cette épineuse question, le *Discours sur notre langue*, écrit sans doute en 1524 vers la fin de sa vie, propose une réflexion significative à cet égard. Machiavel considère, déjà, que la langue constitue un ciment pour une population. S'il ne peut retenir la religion, étant hostile du point de vue politique au christianisme et à la Papauté (Roudier, 2015, pp. 5-16), le Florentin dessine l'importance et la réalité d'une unité par la langue. Dans ce dialogue fictif qu'il mène avec Dante, Machiavel s'oppose à son illustre compatriote sur l'essentiel. Là où l'auteur de la *Divine Comédie* a produit volontairement une langue de Cour, fondamentalement érudite et à destination d'une élite cultivée, celui de la *Mandragore* souligne l'importance d'une langue populaire, parlée et partagée par tous. Contre un Florentin curial, il dégage un Italien populaire, existant à son époque. L'argument central, que nous ne discuterons pas sur le fond : il fait les délices des linguistes, consiste à souligner que si le lexique diffère entre les patois de la Péninsule, les verbes, le « nerf de la langue » sont communs pour l'essentiel.

Ainsi se dégage une patrie possible, lieu où le « si » résonne, sans limite géographique claire mais bien caractérisée par l'usage de variantes inter-compréhensibles de l'Italien.

Machiavel parvient ainsi à creuser la question de l'identité à travers celle de l'altérité. Sans jamais mentionner l'existence de langues différentes dans le Royaume de France, il pousse son analyse de la possibilité d'une unification italienne jusqu'à examiner une éventuelle objection culturelle à son raisonnement et à son objectif. Outre son opposition au christianisme dont il considère, dans les *Discours sur la première décade de Tite-Live*, qu'il empêche les Italiens de retrouver la culture politico-militaire virile des Romains dont la religion est seule idiosyncratique pour les Italiens<sup>3</sup>, Machiavel découvre ainsi un ferment d'unité possible pour son programme politique, aussi « laïc » et matériel que profondément culturel.

L'unification de l'Italie constitue ainsi le programme politique dont le Secrétaire estime que la prise de conscience est nécessaire pour les responsables Florentins. Poussé par les nécessités géopolitiques et géostratégiques issues du déferlement induit par les guerres d'Italie, ce programme est également rendu possible par des circonstances culturelles fondamentales, en particulier par l'existence de variations d'une langue commune qui permet à la *Mandragore* de connaître un succès populaire dès 1518 à travers toute l'Italie. L'utilisation par Machiavel de l'expérience de l'autre lui permet ainsi de cerner certains traits caractéristiques de la modernité politique, en particulier pour la constitution de la question centrale de l'unité politique moderne, ultérieurement nommée « État », voire « État-nation », dont l'existence d'une langue commune forme une marque centrale. Dans notre Europe actuelle où l'on vient de redécouvrir, semble-t-il, que son horizon reste la guerre toujours possible, il n'est pas inutile de rappeler certaines des idées forces de l'inventeur de la science politique moderne, même si le programme qu'il imagina n'est, de toute évidence, désormais plus le nôtre. Nonobstant, il semble illusoire ou à tout le moins dangereux de penser pouvoir éviter les prémices qu'il décèle dans toute situation politique. De notre gré ou par la force des circonstances, nous allons sans doute devoir considérer qu'un machiavélisme raisonnable constitue l'un des outils indispensables, même s'il faut l'adapter, pour maîtriser notre situation et notre avenir.

---

<sup>3</sup> Les vingt premiers chapitres du premier livre des *Discours sur la première décade de Tite-Live* sont sans équivoque sur ce point (Voir également Roudier, 2025).

## Références

### *Œuvres de Machiavel :*

- Machiavel, *Discours sur la première décade de Tite-Live*, Traduction d'Alessandro Fontana et Xavier Tabet, Préface d'Alessandro Fontana, Paris, Gallimard, 2004.
- Machiavel, *Discours sur notre langue*, édition bilingue de Laurent Vallance, Paris, Éditions rue d'Ulm, 2017.
- Machiavel, *Le Prince*, traduction et commentaire de Jean-Louis Fournel et Jean-Claude Zancarini, Paris, PUF, 2000.
- Machiavel, *Œuvres complètes*, Paris, NRF, bibliothèque de la Pléiade, 1952.
- Machiavel, *Toutes les lettres officielles et familières de Machiavel, celles de ses Seigneurs, de ses amis et des siens*, présentées et annotées par Edmond Barincou, préface de Jean Giono, deux tomes, NRF, Paris, 1955.
- Machiavelli, *Il Principe*, A cura di Giorgio Inglese Con un saggio di Federico Chabod, Torino, Einaudi, 1995.
- Machiavelli, *Discorsi sopra la prima deca di Tito Livio*, introduzione di Gennaro Sasso premessa al testo e note di Giorgio Inglese, Milano, Rizzoli, 2010.
- Machiavelli, *Legazioni. Commissarie. Scritti di governo*, Roma, Salerno Editrice, Tome VII (1510-1527) 2011.
- Machiavelli, *Opere*, Torino, Einaudi, Biblioteca della Pléiade, introduction et notes de Corrado Vivanti, 3 tomes, 1999.

### *Autres ouvrages et articles :*

- Botero, G., *De la raison d'État (1589-1598)*, édition, traduction et notes de Pierre Benedittini et Romain Descendre, introduction de Romain Descendre, Paris, Gallimard, 2014.
- Botero, G., *Relationi vniuersali di Giouanni Botero Benese diuise in quattro parti. Nouamente reuiste, corrette, & ampliate dall'istesso auttore. Et aggiuntoui in questa vltima impressione la figurata descrizione intagliata in rame, di tutti i paesi del mondo*, Rome, 1591, Vicence, 1595, Venise, 1596, Brescia, 1599, Turin, 1601.  
[https://archive.org/details/bub\\_gb\\_vxhA3zCi\\_o4C/page/n1/mode/2up](https://archive.org/details/bub_gb_vxhA3zCi_o4C/page/n1/mode/2up)  
consulté le 01/02/2026.
- César, *Guerres. Guerre des Gaules. Guerre civile*. Édition dirigée par Jean-Pierre De Giorgio, Paris, Les Belles Lettres, 2020.
- Chopin, O., Oudet, B., *Renseignement et sécurité*, Paris, Armand Collin, 3ème édition, 2023.
- Fournel, J.L., Zancarini, J.C., *Machiavel, une vie en guerre*, Paris, Passés/Composés, 2020.

- Fournel, J.L., Zancarini, J.C., *Savonarole, l'arme de la parole*, Paris, Passés / composés, 2024.
- Fournel, J.-L., Zancarini, J.-C., *Savonarole, Sermons, écrits politiques et pièces du procès*, Paris, Seuil, 1993.
- Gilbert F., *Machiavel et Guichardin, Politique et histoire à Florence au XVI<sup>e</sup> siècle*, Paris, Seuil, 1996.
- Guidi, A., *Un segretario militante. Politica, diplomazia e armi nel cancelliere Machiavelli*, Bologne, Il Mulino, 2009.
- Levacher, M., *Buffon et ses lecteurs, Les complicités de l'Histoire naturelle*, Paris, Garnier, 2011.
- Marco Polo, *La Description du monde*, [1298] Édition et traduction de Pierre-Yves Badel, Paris, Le Livre de Poche, 2012.
- Montaigne, *Les Essais*, Paris, Gallimard, « la Pléiade », 2007.
- Plancarpin, *Dans l'empire mongol*, [1247] textes rassemblés, présentés et traduits du latin par Thomas Tanase, éditions Anacharsis, Toulouse, 2014.
- Pline l'Ancien, *Histoire naturelle*, Paris, Gallimard, « la Pléiade », 2013.
- Rivière, J.M., *L'expérience de l'autre, les premières missions diplomatiques de Machiavel, Vettori et Guicciardini*, traduction et commentaire de Jean-Marc Rivière, Aix-en-Provence, Presses Universitaires de Provence, 2018.
- Roudier, J., *Machiavel, une biographie, l'apport intellectuel de sa correspondance*, Paris, Cerf, 2019.
- Roudier, J., « Machiavel et le christianisme » in Dignonnet, R., « Figures religieuses » *Mélanges de Science Religieuses*, université Catholique de Lille, 2015, pp. 5-16.
- Roudier, J., *Machiavel par lui-même*, Paris, PUF, 2025.
- Roudier, J., « Performativité et prophétisation : Proposition de lecture des prédications de Savonarole à travers la grille d'analyse d'Austin », in *Mélanges de Science religieuse*, Lille, 1- 2018, pp. 19-34.
- Vasco de Gama. *Le premier voyage 1497-1499*, La relation attribuée à Álvaro Velho & les lettres des marchands florentins présentées et traduites par Paul Teyssier, Paris, Chandeigne, 2017.
- Verrier, F., *Voyages en Afrique noire d'Alvise Ca' da Mosto (1455 & 1456)*, Relations traduites de l'italien & présentées par Frédérique Verrier, Paris, Chandaigne / Unesco, 2003.

## **Media Discourse and Political Polarization in the Coverage of the Valencia DANA. A Comparative Analysis of *Público* and *Okdiario***

**Cristina LÓPEZ GONZÁLEZ<sup>1</sup>, Carolina BEIRO GARCIA<sup>2</sup>,  
Nuria FIGUEIRA MARQUEZ<sup>3</sup>**

**Abstract.** This study examines how ideological bias in digital media contributes to political polarization through comparative analysis of news coverage by comparing the headlines published by *Público* (left-leaning) and *Okdiario* (right-leaning), two major Spanish online newspapers, regarding the Valencia DANA disaster that occurred in October 2024. Using content analysis, we analyzed 177 political mentions in the media outlets, examining both manifest content (frequency of political figures and parties mentioned) and latent content (tone and framing). The findings reveal significant differences in how the outlets framed the same crisis event: *Público* predominantly criticized the Partido Popular and regional government, while *Okdiario* focused criticism on the PSOE and central government. The results show that media outlets do not merely report events neutrally but actively construct polarized narratives reflecting their editorial stances, amplifying political divisions, particularly during crises, through selective framing and emphasis.

**Keywords:** framing, media polarization, crisis communication

---

<sup>1</sup> Centro de Estudios Superiores Universitarios de Galicia (CESUGA), alu.168365@usj.es

<sup>2</sup> Centro de Estudios Superiores Universitarios de Galicia (CESUGA), alu.166322@usj.es

<sup>3</sup> Centro de Estudios Superiores Universitarios de Galicia (CESUGA), alu.168372@usj.es



## Introduction

In recent years, political polarization has become one of the most defining characteristics of contemporary societies. This polarization is not only reflected in political institutions and public debates but also in the way media outlets construct and disseminate news. Digital media, in particular, play a crucial role in shaping public opinion due to their immediacy, wide reach, and strong ideological positioning (Happer & Philo, 2013).

The DANA that affected Valencia in late October 2024 represents a particularly relevant case for the study of media discourse and political polarization. Beyond its severe human and material consequences, the event quickly became a central topic of political confrontation, with different actors blaming one another for the management of the crisis. As a result, media coverage of the DANA extended far beyond the reporting of a natural disaster and became a politically charged narrative.

This research focuses on how digital media outlets with different ideological orientations framed and narrated the same event. Specifically, it compares the coverage of the Valencia DANA by *Público*, a left-leaning digital newspaper, and *Okdiario*, a right-leaning online outlet. (Monroy-Trujillo & Padilla-Castillo, 2025) By analyzing news articles published during the two weeks following the natural disaster, this study aims to identify differences in tone, language, framing, and the portrayal of political actors.

The choice of this topic is motivated by the increasing influence of media discourse on social behavior and political attitudes. In highly polarized contexts, media narratives can reinforce ideological divisions, shape perceptions of responsibility, and influence how citizens interpret crises. The DANA in Valencia provides a clear example of how a humanitarian and environmental tragedy can be transformed into a political and media battleground.

## Political and media polarization

Political polarization refers to the growing ideological distance between political positions, parties, and social groups, resulting in increasingly divided public opinions and reduced consensus. In polarized contexts, political actors and citizens tend to adopt more extreme positions, often perceiving opposing views as illegitimate or threatening. This phenomenon has intensified in recent decades due to globalization, economic inequality, and the rise of digital communication technologies (Boxell et al., 2017).

Iyengar, Sood, and Lelkes (2012) distinguish between ideological polarization and affective polarization. While ideological polarization concerns differences in political beliefs and policy preferences, affective polarization refers to the emotional dimension of politics, in which individuals develop strong negative feelings toward opposing political groups. This affective dimension is especially relevant in media studies, as emotions play a central role in news consumption and interpretation.

Political polarization has significant consequences for democratic societies. It can undermine trust in institutions, weaken social cohesion, and increase political conflict. In moments of crisis, such as natural disasters or emergencies, polarization may intensify debates over responsibility, competence, and legitimacy, shifting the focus from collective problem-solving to political confrontation.

In the context of media communication, political polarization is closely linked to the role of news outlets as mediators between political actors and the public. (Entman, 2010) Media do not merely transmit information neutrally; they select, frame, and interpret events, often reflecting specific ideological orientations. As a result, polarized political environments tend to be accompanied by polarized media systems, in which audiences consume news that aligns with their preexisting beliefs.

Media polarization refers to the tendency of news outlets to present information in a way that reflects particular political or ideological perspectives. In polarized media systems, audiences are often exposed to content that reinforces their preexisting beliefs, while alternative viewpoints may be marginalized or framed negatively. This selective representation can amplify political divisions and shape public opinion in subtle but significant ways.

According to Stroud (2011), ideological bias in the media can be observed through both manifest and latent content. Manifest content includes explicit facts, figures, and events reported in the news, whereas latent content refers to the implicit messages, tone, and framing that communicate approval or disapproval of certain actors or policies. Analyzing both types of content is essential for understanding how media narratives influence perceptions and political attitudes.

In digital media, polarization is particularly pronounced due to algorithmic filtering, social media sharing, and the rapid dissemination of information. Online platforms allow users to select news sources that align with their ideological preferences, creating “echo chambers” in which polarized

messages are amplified. This phenomenon makes the study of media bias in digital news especially relevant in the context of politically charged events, such as the coverage of the Valencia DANA.

The ideological orientation of media outlets can influence several aspects of news reporting, including the selection of events, the language used in headlines, the portrayal of political figures, and the framing of causes and responsibilities. By comparing news coverage from left-leaning (*Público*) and right-leaning (*Okdiario*) outlets, this study aims to examine the mechanisms through which ideological bias shapes public understanding of a single event.

### **Crisis communication**

Crisis communication refers to the strategies and practices used by organizations, institutions, and media outlets to manage information and public perception during emergencies or critical events. Effective communication during a crisis is essential to inform the public, coordinate responses, and maintain trust. In the context of natural disasters, such as floods, storms, or the DANA in Valencia, media coverage plays a central role in shaping public understanding of both the event and the actions taken by authorities.

Coombs (2015) emphasizes that media not only report facts during crises but also frame narratives that can influence public sentiment and political accountability. The way a crisis is communicated – through choice of words, tone, and emphasis – can either foster solidarity and cooperation or generate fear, confusion, and political controversy. For instance, focusing on human stories and rescue efforts tends to build empathy, whereas highlighting political disputes can intensify polarization.

In highly polarized media environments, crisis communication often becomes entangled with political agendas. News outlets may select which aspects of a disaster to emphasize, attribute responsibility to specific actors, or use emotionally charged language to support ideological narratives. This creates an additional layer of complexity, as citizens receive not only information about the crisis itself but also signals about who is “responsible” or “competent” in managing it.

In the case of the Valencia DANA, examining how different news outlets reported the event allows researchers to analyze the intersection of natural disasters, media framing, and political polarization, providing insight into the broader dynamics of media influence in contemporary society.

## **Framing the crises**

Framing theory suggests that the way information is presented—its structure, emphasis, and context—can significantly influence how audiences interpret events. Frames highlight certain aspects of reality while downplaying or ignoring others, shaping perceptions, opinions, and even emotional responses. According to Entman (1993), framing involves selection and salience: selecting some aspects of perceived reality and making them more noticeable, meaningful, or memorable in communication.

In the context of digital media coverage of crises, framing plays a crucial role in how political responsibility, urgency, and the social impact of an event are communicated (Entman, 1993; Cho & Gower, 2009). For example, emphasizing government failures or political conflicts in a natural disaster may generate outrage and blame, while highlighting rescue efforts and solidarity may foster empathy and trust (Semetko & Valkenburg, 2000). Frames are not neutral; they reflect both the ideological orientation of the media and the broader political environment (Entman, 2010). Studies have examined media coverage of specific crises to understand polarization effects. For example, research on hurricane coverage in the United States by Boykoff and Boykoff (2004) found that left-leaning and right-leaning outlets reported the same events with differing emphases on causes, responsibility, and policy implications. Similarly, studies on European floods and wildfires (Schweiger, 2019) indicate that media with divergent ideological perspectives tend to highlight different actors, causes, and solutions, reflecting broader political divides.

Research on crisis communication in digital contexts stresses the relevance of online platforms as arenas for political contestation. Digital media not only report events but actively shape discourse, as users share, comment, and interact with content, potentially reinforcing polarization. Tandoc and Vos (2016) emphasize that social media amplifies ideological framing, making the analysis of online news coverage essential for understanding contemporary media influence.

Specific research on Spanish media confirms that political polarization influences reporting on national and international events. According to a study by Sampedro (2020), newspapers such as *El País* and *ABC* consistently display ideological framing in both editorial and news articles, affecting the

tone, word choice, and presentation of political figures. These patterns are comparable to the case of *Público* and *Okdiario*, in which left-leaning and right-leaning outlets frame the same events differently, emphasizing elements that appeal to their respective audiences. This demonstrates that media polarization is not merely theoretical but has observable effects in practice.

By comparing the coverage of the Valencia DANA in *Público* and *Okdiario*, this research formulates two research questions:

RQ1: How ideological bias shapes news narratives in the context of a natural disaster?

RQ2: How the particular coverage contributes to political polarization in Spain?

### **Advantages and limitations**

This study presents several advantages that contribute to the understanding of media discourse and political polarization. First, by comparing two ideologically opposed digital news outlets, *Público* and *Okdiario*, the research allows for a direct examination of how political bias shapes news narratives. This comparative approach provides insights into differences in framing, language use, and portrayal of political actors, highlighting the mechanisms through which polarization manifests in media coverage.

The use of content analysis as a research method to identify both manifest and latent content of the texts (counting mentions of political figures, parties, and events combined with an assessment of tone, sentiment, and semantic choices enables a comprehensive understanding of media narratives). This dual approach ensures that the study captures not only the frequency of information but also the qualitative nuances that influence audience perception.

Third, the focus on a recent and highly relevant event – the Valencia DANA of October 2024 – enhances the study's practical significance. By analyzing a current crisis that attracted intense political and media attention, the research addresses a real-world phenomenon and provides timely insights into the interaction between media, politics, and society.

Despite these advantages, the study also has several limitations that must be acknowledged. The sample is limited to two digital news outlets, which may not fully represent the broader Spanish media landscape. Including

additional newspapers, television channels, or social media platforms could offer a more complete picture of media polarization. Additionally, the analysis covers a relatively short time frame – the two weeks following the DANA – which may limit the ability to observe longer-term media trends or evolving narratives.

Finally, the focus on Spanish-language media may limit the generalizability of the findings to other countries or linguistic contexts, although the theoretical framework remains applicable across settings.

## **Research design**

For this research analysis, we developed a comparison regarding political polarization and influence in mass media. In this case, the media chosen for the research study are *Público* and *Okdiario*, two recognized online newspapers from Spain. They were both selected because of their typically opposite editorial tone, which makes them ideal for comparison, *Público* being more progressive, while *Okdiario* being more conservative), and also because of their free, complete, and easy access to their past published content through the archives. The accessibility of the content that is presented here makes it reliable and simply replicable. As already stated, the topic of the articles used for the research revolves around politics within the climatic disaster called DANA that occurred in the Spanish province of Valencia in October 2024.

The research method used for this study is content analysis, including both the manifest and the latent layers of the texts. Manifest content includes measurable aspects such as the frequency of mentions of political figures or parties, while latent content captures tone, sentiment, and connotations. Studies using this dual approach (Neuendorf, 2017) have shown that combining quantitative and qualitative measures provides a more complete understanding of media influence and audience interpretation.

The sample selected for the analysis has been the headlines of every article related to the topic published by each newspaper between the day of the incident and two weeks later, ranging from October 29, 2024 to November 12, 2024. The articles for *Público* can be found online on their website under the tag named 'DANA'. On the other hand, for *Okdiario*, the information is found on their page under the tag of 'Floodings'. Not all of the

headlines that fit those requirements were analyzed, but only those referring to politics. The analysis grid focused on two different aspects observed in the selected headlines. First, the manifest content: which political parties and/or personalities were mentioned and their frequency. Second, the latent content: what was the tone in which those mentioned were referred to. In other words, what was the vocabulary used to portray them: positive, negative, or more neutral.

The procedure for the research was conducted completely manually and was as follows: First, we looked for the information regarding this incident in each of the publications and found the most suitable tag that should include the data needed for the study. Afterward, we separated the articles published during the defined time frame and started to pick from the whole the headlines that referred to politics. Then, every time a political personality or party was mentioned, it was highlighted in the text. Different colors were used for this step, symbolically, to differentiate which party each person was part of. Next, the mentions highlighted were counted for each color but also for each newspaper. That is, for example, how many times words were highlighted in blue (referring to Partido Popular) in the *Público* headlines and how many were in *Okdiario* headlines. After that, the data was recounted and divided into positive, negative, or neutral tone used for each party. Finally, the mentions for each political group were transformed into the percentages relative to the total of political mentions made in each newspaper.

## Research results

*Público* (n=156 political mentions) has a very rich variety of news and articles published along the time frame of this study. From opinion genres to up-to-the-minute news, updating constantly with new information that was known. Included in the headlines of the articles selected by the criteria already explained are 156 political mentions of parties, groups, and personalities.

**Partido Popular (PP):** There are 81 mentions of this group, mostly naming Carlos Mazón, President of the Generalitat Valenciana from July 17, 2023, until December 2, 2025; Alberto Núñez Feijóo, president of the party;

the party itself; and the government of the Generalitat Valenciana. The vast majority of these mentions have a negative undertone, criticizing the management of the issue by the regional government, mostly Carlos Mazón and the emergency counselor. Out of the 81 mentions, 66 are classified as negative (e.g.: “Mazón llegó con dos horas de retraso a la reunión de emergencia por la DANA en València” / Mazón was two hours late to the emergency meeting for the DANA in Valencia), 13 as neutral (e.g.: “La Generalitat desmiente que maneje ninguna cifra concreta de personas desaparecidas por el momento” / The Generalitat denies that they are working with a specific number of disappeared people at this moment), and 2 as positive (e.g.: “Mazón pide que se incorporen 5.000 militares más a las tareas de emergencia” / Mazón asks for the incorporation of 5,000 more soldiers for the emergency work) toward the party. The total mentions of Partido Popular represent 51.92% of the Público political mentions in the sample.

**Partido Socialista Obrero Español (PSOE):** The total mentions were 54 and they represented 34.62% of the total mentions in Público. Most of them name Pedro Sánchez, President of Spain and General Secretary of the party, the Spanish Government, and Teresa Ribera, third Vice President of the Government. It is also noteworthy to mention the references to Fernando Grande-Marlaska, Minister of Internal Affairs, and Óscar Puente, Minister of Transport and Sustainable Mobility. 34 out of the 54 mentions to this group were positive, defining the Government management as helpful, available, and reliable with the Generalitat (e.g.: “El Gobierno evita la confrontación directa con el PP y prepara un plan de ayudas inminente para los afectados de la DANA” / The Government avoids direct confrontation with PP and prepares an imminent help plan for those affected by the DANA). The other 20 mentions are divided between 13 neutral ones, just providing information (e.g.: “Sánchez hará una declaración institucional sobre la DANA a las 10.30 horas desde Moncloa” / Sánchez will make an institutional declaration regarding the DANA at 10:30 from Moncloa), and 7 negative ones (e.g.: “Ojalá Óscar Puente haciéndose cargo del armario de mis 'tuppers': las redes se llenan de memes con la labor del ministro tras la DANA” / “I wish Óscar Puente was taking care of my tupper drawer”: social media fills up with memes about the minister's work after the DANA).

**Far right parties and politicians:** There are 8 mentions of this group, all classified as having a negative undertone. This includes the political party Vox, its President Santiago Abascal, and neo-fascist groups (e.g.: “La ultraderecha

vuelve a utilizar los tribunales para sacar rédito político con la DANA” / The far right uses the courts once again to get political yield from the DANA). The total mentions of this group are 5.13% of the mentions in Público.

**The King and the Queen of Spain:** They are mentioned 7 times, all regarding their visit to the ground zero of the disaster, alongside Pedro Sánchez and Carlos Mazón. Six of these mentions are classified as neutral (e.g.: “Sánchez, evacuado de Paiporta tras una agresión en la visita con los reyes y Mazón” / Sánchez evacuated from Paiporta after an aggression during the visit with the kings and Mazón), and one as negative (e.g.: “Entre Mazón en chaleco y el Borbón en traje de faena, todos salvados”: el papelón de los políticos con la DANA, explicado en tuits” / Between Mazón in a vest and the Bourbon in work clothes, all saved: the spectacle of politicians with the DANA, explained in tweets), and they represent 4.49% of the total.

**Podemos and Sumar:** Each of these two political parties is mentioned twice in the Público article headlines, all positive. In both cases, the party itself is named once and the other mention is a member of it. For Sumar, that person is Yolanda Díaz, founder of the party and second Vice President of the Government of Spain. Each represents 1.28% of the mentions. (e.g.: “Podemos y Compromís quieren que el Congreso investigue las responsabilidades políticas de la tragedia de la DANA” / Podemos and Compromís want the Congress to investigate the political responsibilities in the DANA tragedy) (e.g.: “Yolanda Díaz protegerá el teletrabajo y ausentarse del empleo a los afectados por la DANA y ofrece permisos retribuidos” / Yolanda Díaz will protect remote working and abstention from work for those affected by the DANA and offers rewarded permits)

**Others:** The Valencian political party Compromís and Gabriel Rufián, representative of Esquerra Republicana de Catalunya in the National Congress, also receive a mention each. Both have a positive undertone and together represent another 1.28% of the mentions. (e.g.: “Podemos y Compromís quieren que el Congreso investigue las responsabilidades políticas de la tragedia de la DANA” / Podemos and Compromís want the Congress to investigate the political responsibilities in the DANA tragedy) (e.g.: “Rufián responde al PP recordando sus grandes embustes: “Mintieron con Irak, las preferentes, el rescate a la banca, el Prestige, el Yak-42...” “ / Rufián answers PP recalling their big lies: “They lied with Iraq, the preferentials, the bank rescue, the Prestige, the Yak-42...”)

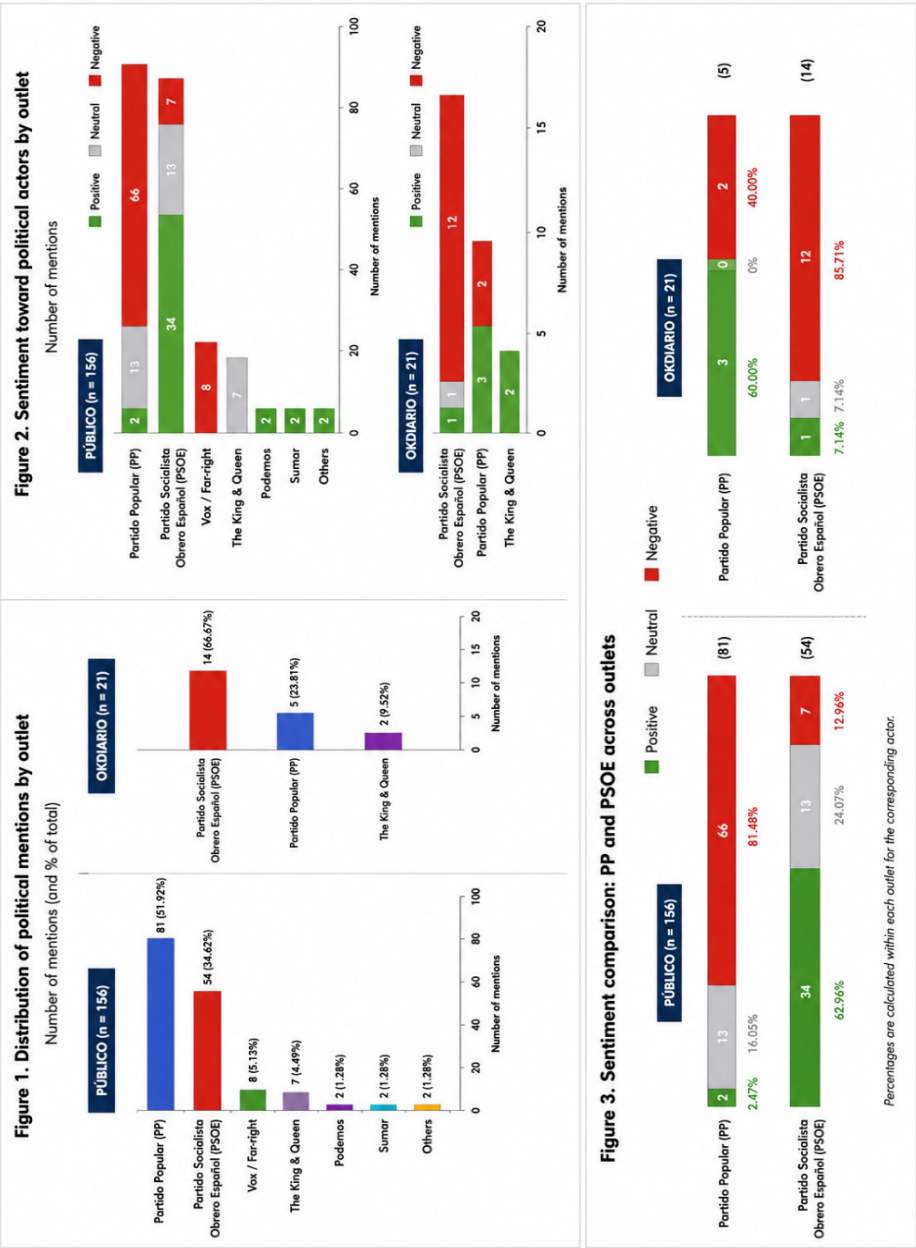


Figure 1. Distribution of political mentions by outlet  
Figure 2. Sentiment toward political actors by outlet  
Figure 3. Sentiment comparison – PP and PSOE across outlets

Okdiario (n=21 political mentions) had a more limited number of articles about the issue of this study, especially after reducing the headlines selected to those related to politics. Those that meet the criteria explained before sum up to 21 political mentions. The groups included in them are the following:

**PSOE:** It is mentioned 14 times, reaching 66.67% of the total political mentions that appear in Okdiario's headlines. Again, Pedro Sánchez, the minister Teresa Ribera, and the party itself are the most named. The vast majority of those, exactly 12 mentions, have a negative undertone (e.g.: “El Estado se gasta 2 millones al año en una red para detectar riadas en Valencia que no evitó la tragedia” / The State spends 2 million per year on a system for detecting floodings in Valencia that did not prevent the tragedy). Another mention is more neutral (e.g.: “Feijóo dice que España está ante una “emergencia nacional” que requiere “todos los recursos del Estado”” / Feijóo says Spain is facing a “national emergency” that requires “all of the State resources”), and the remaining one is the only positive one (e.g.: “Sánchez propuso una declaración conjunta “contra los violentos” en Paiporta pero Mazón la rechazó” / Sánchez proposed a joint statement “against the violent” in Paiporta but Mazón turned it down).

**PP:** The total of mentions they receive from Okdiario is 5, representing 23.81%. Carlos Mazón, Alberto Núñez Feijóo, the Generalitat, the counselor Nuria Montes, and the party itself are the most named elements. Three of the mentions are positive (e.g.: “El PP no es el PSOE: Mazón debe destituir ya a la consejera” / PP is not PSOE: Mazón needs to dismiss the counselor already (this one actually includes both positive and negative tones)), and the remaining two are negative (e.g.: “Sánchez propuso una declaración conjunta “contra los violentos” en Paiporta pero Mazón la rechazó” / Sánchez proposed a joint statement “against the violent” in Paiporta but Mazón turned it down).

**The King and the Queen:** In the case of Okdiario, they are named in 9.52% of the total with two mentions. Both of those mentions have a positive tone, picturing them as key figures for the stability of the country. (e.g.: “Cuando el Rey era y fue el Estado” / When the King was and is the State).

These results show very clearly the political polarization of mass media and how the words we use to tell the same story can vastly change how it is perceived by the public. In a case like the disaster of the DANA in

Valencia, the controversy regarding who is to blame still surrounds the Generalitat Valenciana and the National Government. If somebody read *Okdiario* to inform themselves about the event, they would have probably thought Pedro Sánchez and his government were responsible for the mismanagement. However, if they read *Público*, they would have been critical of Mazón's way of acting. The results of the study confirm the monarchic and conservative posture of *Okdiario* and the progressive and clearly left-oriented style of *Público*.

## Discussion and conclusions

This research set out to examine how media discourse contributes to political polarization through the coverage of the same political event by ideologically opposed digital media outlets. By focusing on the coverage of the DANA climatic disaster that occurred in Valencia in October 2024 and comparing the headlines published by *Público* and *Okdiario*, this study aimed to understand how political ideology shapes media narratives, framing strategies, and the portrayal of political actors during moments of crisis.

The results of the content analysis clearly show that media outlets do not merely inform about political events but actively construct narratives that reflect their editorial positions. The comparison between *Público* and *Okdiario* reveals significant differences in both the quantity and tone of political mentions. *Público* showed a much higher number of political references, with a strong focus on the Partido Popular and the Generalitat Valenciana, predominantly framed in a negative tone. This framing emphasized responsibility, mismanagement, and political accountability, particularly targeting Carlos Mazón and the regional government. In contrast, *Okdiario* presented a considerably smaller sample of political mentions, concentrating mainly on the PSOE and portraying it largely in a negative light, while framing the Partido Popular and the monarchy more positively.

These findings align with existing literature on media polarization and ideological bias, which suggests that news outlets often act as political actors rather than neutral observers. As discussed in the theoretical framework, media framing theory explains how the selection of certain words, tones, and actors can guide audience interpretation. In this case, both newspapers framed the DANA not only as a natural disaster but also as a political conflict, reinforcing narratives that align with their ideological orientations. This

supports previous research indicating that crises are often instrumentalized by media to reinforce preexisting political positions and to allocate blame or legitimacy selectively.

The research questions posed at the beginning of the study can now be addressed. Firstly, political actors and institutions play a central role in shaping media discourse, not only through their actions but also through how media choose to represent them. The prominence of certain figures, such as Pedro Sánchez or Carlos Mazón, and the repeated association of these actors with either competence or failure, illustrates how media discourse contributes to constructing political legitimacy or delegitimization. The results show that political responsibility during the DANA was framed very differently depending on the outlet, reinforcing the idea that media narratives are deeply intertwined with political power dynamics.

Secondly, the polarized framing observed in both outlets may contribute to reinforcing ideological divisions among audiences. Readers consuming only one of these media sources are likely exposed to a one-sided narrative, which may strengthen political attitudes, reduce empathy toward opposing viewpoints, and increase mistrust in political institutions associated with the opposing ideology. This phenomenon is particularly relevant in the context of digital media ecosystems, in which algorithmic personalization often amplifies selective exposure.

The study confirms that the influence of political ideology varies significantly between left-leaning and right-leaning outlets. *Público* adopted a critical stance toward conservative actors while highlighting positive actions by the central government, whereas *Okdiario* emphasized criticism of the PSOE and minimized or positively framed conservative responsibility. This asymmetry illustrates how ideological positioning influences not only what is reported but also how frequently and in what tone political actors appear in media discourse.

Regarding the question of misinformation or strategic omission, while this study did not directly analyze factual inaccuracies, the selective emphasis and framing observed may contribute to distorted perceptions of reality. By prioritizing certain political interpretations over others, media outlets may indirectly contribute to the spread of partial or biased narratives, even without explicitly publishing false information. This reinforces concerns raised in previous studies about the role of media in shaping political reality rather than merely reflecting it.

Finally, concerning whether media discourse anticipates future political developments, the findings suggest that media narratives may indeed foreshadow political conflicts by reinforcing blame, accountability, and legitimacy narratives early on. The strong polarization observed in the immediate aftermath of the DANA indicates that media coverage can contribute to shaping long-term political debates and public memory of events, influencing how responsibility is assigned beyond the immediate crisis.

In conclusion, this research highlights the powerful role of digital media in shaping political discourse and reinforcing polarization, especially during moments of crisis. By comparing two ideologically opposed Spanish online newspapers, the study demonstrates how the same event can be narrated in fundamentally different ways, depending on editorial orientation. These findings are relevant for media studies as they emphasize the need for critical media literacy and awareness of ideological framing in digital news consumption. Understanding how media narratives are constructed is essential for fostering a more informed and reflective public, capable of navigating an increasingly polarized media environment.

## References

- Boykoff, M. T., & Boykoff, J. M. (2004). Bias in the media: How mainstream news coverage reflects corporate interests. *Global Environmental Change*, 14(3), 205–214.
- Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). Is the internet causing political polarization? Evidence from demographics (No. w23258). *National Bureau of Economic Research*.
- Cho, S. H., & Gower, K. K. (2009). How do the news media frame crises? A content analysis of crisis news coverage. *Public Relations Review*, 35(2), 107–112.
- Coombs, W. T. (2015). *Ongoing crisis communication: Planning, managing, and responding* (4th ed.). Sage Publications.
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58.
- Entman, R. M. (2010). Media framing biases and political power: Explaining slant in news of Campaign 2008. *Journalism*, 11(4), 389–408.
- Happer, C., & Philo, G. (2013). “The Role of the Media in the Construction of Public Belief and Social Change.” *Journal of Social and Political Psychology*, 1(1), 321–336. DOI: <https://doi.org/10.5964/jspp.v1i1.96> [jspp.psychopen.eu](http://jspp.psychopen.eu)

- Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, 76(3), 405–431.
- Monroy-Trujillo, A., & Padilla-Castillo, G. (2025). Media Bias in Immigration Reporting: A Comparative Study of Spanish Newspapers' Source Usage. *Journalism and Media*, 6(4), 160.
- Neuendorf, K. A. (2017). *The content analysis guidebook* (2nd ed.). Sage Publications.
- Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin Press.
- Sampedro, V. (2020). Dominating strategy: Spain's media system in a comparative perspective. *Media, Culture & Society*, 42(3), 389–407.
- Semetko, H. A., & Valkenburg, P. M. (2000). Framing European Politics: A Content Analysis of Press and Television News. *Journal of Communication*, 50(2), 93–109.
- Schweiger, W. (2019). *Digital erosion: Rethinking the internet in the age of algorithmic distortions*. Springer.
- Stroud, N. J. (2011). Niche news and political participation: How consuming intentionally and unintentionally political news affects civic engagement. *Journalism*, 12(3), 311–327.
- Tandoc Jr, E. C., & Vos, T. P. (2016). The journalist is a citizen now: Social media and journalistic norms. *Digital Journalism*, 4(7), 904–923.

## Who Are We in the World of GIFs? - Mapping GIF Users

Erika-Zita CSÓKA-ANTAL<sup>1</sup>

**Abstract.** In online communication, the effective transmission of emotions, reactions, and states is becoming increasingly important, and creative solutions such as emojis, GIFs, and memes, which are capable of reflecting an individual's feelings, reactions, or even digital identity. Since their return, reaction GIFs have become increasingly popular as a means of expressing nonverbal cues in the online space. These animated reaction GIFs are not only capable of expressing a single emotion or set of gestures, but often carry pop culture references, knowledge from meme culture, or even inside jokes, making them multifaceted and adding color to online communication. The use and interpretation of GIFs can be considered a complex process that involves different skills, motivations, background knowledge, or even literacy, and based on these, similar to previous media such as television or even radio, we can identify different user groups for GIFs.

The present research is based on the study by Church, King et al. *“Relating, searching, and referencing: Assessing the appeal of using GIFs to Communicate”*, which identifies three main categories of GIF users: enthusiasts, searchers, and referentialists. These general categories help us understand the purposes and motivations behind people's use of GIFs in online communication. The purpose of the research is to examine GIF users among Generation Z university students. More specifically, it examines the frequency with which these “pure” categories appear, the possible transitions between them, and whether there are any additional categories beyond those mentioned. The methodology is based on a survey questionnaire, the main purpose of which is to collect structured data on views and attitudes related to GIFs, using the categorizations and statements of the underlying research study.

**Keywords:** GIF, Generation Z, user, digital identity, digital communication

---

<sup>1</sup> PhD Student, Babeş-Bolyai University, erika.csoka@ubbcluj.ro



## GIF (Graphic Interchange Format)

GIF, or Graphic Interchange Format, is a file type that was primarily used as a web design element for unfinished websites in the early days of the internet. At the time, it was useful because it displayed images efficiently and quickly in an online environment. The original concept was to allow graphic elements to be shared easily, which contributed significantly to the rise of interactive and interactive content in the online world (Eppink, 2014). After that, the popularity of GIFs declined until 2006, when they returned to public consciousness on Tumblr, serving an illustrative and aesthetic function in various posts, adding to various Tumblr blog posts (Coen, 2016). From then on, the popularity of GIFs in the digital environment has only grown. To the point that today, GIFs are not considered a file format, but are known as looped animations that are an essential part of online communication.

According to Ágnes Veszelszki's definition, "GIFs are short moving images that can be sent in response to various conversational situations in digital communication, serve to express emotions, can be considered a more intense form of emotional expression than emoticons, mostly show scenes extracted from well-known films, one of their main components is humor, and in stylistic terms, they use exaggeration as a device" (Veszelszki, 2015:79). In his work *The brief history of GIFs*, American author Jason Eppink states that sending a GIF is a reaction that shows HIFW how I feel when, or MRW, what my reaction is in a given situation (Eppink, 2014: 303).

The success of GIFs lies not only in their ability to capture iconic moments in pop culture, but also in their ability to effectively express emotions (Highfield, 2017). GIFs are able to transmit emotions more intensely than other screen-based emoticons. While an emoticon usually expresses facial expressions, GIFs are capable of displaying gestures, body movements, and even eye movements in addition to facial expressions, and the dynamic sequence of movements presents the given state or reaction in a more intense and illustrative way. The looped motion allows the GIF to be more emphatic and memorable, highlighting important details. Due to their dynamic nature, GIFs also introduce the dimension of temporality into the repertoire of emotional expression (Veszelszki, 2015).

The looped motion of GIFs removes the sense of finality and transience from the act of communication. In addition to expressing emotions, an important feature of GIFs is their ambiguity. Unlike emojis, GIFs do not have

a specific meaning, which allows for free interpretation. In addition, the repetitive motion also carries with it the possibility of open and continuous interpretation (Highfield, 2017). This enable the process of decontextualization, as a result of which the GIF stands on its own, even if the user is not familiar with the original context of the GIF, the user can easily recognize and interpret the message they communicate, as the sequence of movements and their ambiguity allow them to be interpreted and used independently in a digital environment.

Recently, the role and importance of GIFs in internet communication has steadily increased. Their function today differs from their original purpose, and they are now suitable for expressing more complex feelings and states (Eppink, 2014). The success of GIFs depended in large part on their adaptability. These moving images were able to keep up with changes in the Internet environment and continuously developed and expanded their functionality. Among these developments, the emergence of GIF galleries in 2013, such as Giphy and Tenor, was of particular importance, as they allowed users to easily browse and search for GIFs based on keywords. However, the real breakthrough came in 2016 when these GIF search engine galleries were integrated into social media platforms and various communication applications. This step significantly improved the accessibility and usability of GIFs, promoting the development of user-friendly GIF usage. Currently, GIFs can be found on almost every online platform, whether it's social media sites like Facebook or X, or even communication apps like WhatsApp and Messenger, and they even appear in specific apps such as dating apps like Tinder (Meier, 2018). This general presence means that GIFs are an integral part of users' online experience, closely intertwined with their everyday digital communication and entertainment. Digitalization has not only changed the forms and tools of communication, but also transformed language itself, creating a unique linguistic system that many have called net language (Curry, 2008), digilect (Veszelszki, 2010), or written chat (Wirth, 2000). In this new linguistic environment, visual emoticons such as GIFs also have their place. As a result, GIFs can even function as a language, as they have been able to integrate into the keyboards of our digital devices and carry specific cultural codes in digital communication, serving as a kind of digital emoticon. At the same time, they embody one of the most important characteristics of digital language, namely technological determinism, which means that the possibilities and limitations of the technology or medium determine linguistic

behavior, and therefore it must constantly keep pace with innovations (Veszelszki, 2017). This means that digital language is constantly changing and adapting to new technological possibilities and developments.

Moreover, GIFs offer users the opportunity to express their feelings and moods in a personalized way in the online space. In this way, they play an important role in both the creation and shaping of users' online identities (Wang et al., 2024). Choosing a GIF not only expresses emotions, but can also reflect the user's unique style, interests, and even cultural references. GIFs allow users to express themselves in a unique way and identify with online communities, enhancing their online presence and connections.

The use of GIFs has become particularly important in expressing emotions and nonverbal cues, similar to other tools spreading on the internet, such as emojis, stickers, and memes. GIFs are able to go beyond emojis because they offer a more dynamic form of expression. While emojis can only depict a limited number of facial expressions and objects, GIFs show entire sequences of movements (Veszelszki, 2015). These captured animations not only depict facial expressions realistically, but also body language, gestures, and other signals. In addition, their repetitive, looped movement often serves to emphasize the most memorable parts of scenes, thus GIFs offer a much richer set of communication tools that can be used to express emotions and thoughts much more accurately and colorfully than simpler, static emojis.

At the same time, GIFs also show similarities to internet memes. There is often a GIF version of a meme, or we can talk about memed GIFs, but it is important to note that although both are part of internet folklore and reflect the creativity of users (Wagener, 2021). While memes focus on recreation, GIFs focus on reuse. There are also similarities in terms of their success factors, but while internet memes can convey humor and related emotions and states, such as sarcasm, cynicism, etc., GIFs have a wider palette and, in addition to basic emotions, are capable of expressing feelings such as empathy, closeness, sympathy, etc.

## **Previous research**

With the growing popularity of GIFs, researchers are becoming more interested in various aspects of these small moving images. Research on GIFs mostly examines their technical properties, production, and website

compatibility. However, qualitative research also covers their interpretation (Brubaker, 2017), methods for finding the perfect GIF (Jiang, 2018), and their cultural significance (Highfield, 2017). From the perspective of communication science, it is also interesting to examine the communicative properties of GIFs and how they return to the public consciousness, even at the lowest point of their popularity. Because of their complexity, the interpretation of GIFs cannot be separated from their use. The relationship between the interpretation and use of GIFs is analyzed in a study by Church, King, and colleagues entitled “*Relating, searching, and referencing: Assessing the appeal of using GIFs to communicate.*” The study's main goal was to examine GIFs from a communication perspective and understand user behavior. The research identified three main groups: GIF enthusiasts, GIF searches, and GIF referentialists. These groups were categorized based on factors such as the technical characteristics of GIFs, the search process, emotional expression, cultural significance, expression of humor, linguistic significance, and the importance of relationships.

The table below summarizes the typology of GIF users and their main motivations:

**Table 1.** GIF user typologies and their most important characteristics  
(Own editing)

| GIF characteristics                | Enthusiasts | Searchers | Referentialists |
|------------------------------------|-------------|-----------|-----------------|
| <i>Technical properties</i>        |             | x         | x               |
| <i>Language</i>                    | x           |           |                 |
| <i>Expression of affect</i>        | x           |           |                 |
| <i>Culture</i>                     |             |           | x               |
| <i>Humor</i>                       | x           | x         |                 |
| <i>Searching</i>                   | x           | x         |                 |
| <i>Importance of relationships</i> | x           |           | x               |

The research used Q methodology from a methodological point of view, using the usage and satisfaction model as a starting point. This approach allowed them to conduct the research in two phases. In the first phase, the researchers measured attitudes toward GIFs. In this part, participants were asked to rank 32 statements using a survey. The purpose of this was to identify different views and attitudes towards GIFs based on attitudes, thereby enabling the separation of larger categories. Statistical

analysis was used to associate different statements with each user group. In the multi-scale attitude measurement, the statements and values with high positive values from the 32 statements helped to outline the most important motivations and characteristics, and similarly, the statements with predominantly or significantly negative values highlighted the biggest differences between the three categories.

In the second phase, the participants were interviewed, which helped to create and more precisely define the different categories. This methodological approach allowed the researchers to examine attitudes and user behavior related to GIFs in a comprehensive and detailed way, as well as the motivations and considerations that participants used to categorize GIFs.

### **Characteristics of the different groups, based on the research results:**

1. **Enthusiasts:** Enthusiasts mainly see the emotional richness, humor, and communicative power of GIFs. For them, GIFs are expressive tools that can transmit more emotion than emojis. They do not focus on searching, but rather on obtaining the desired result, for example making their friends laugh. These people primarily use GIFs in their social interactions, as they find them to be highly effective on an emotional and communicative level. For them, GIFs serve to express emotions, especially humor, helping to strengthen relationships and add dynamism to conversations. The visual nature of GIFs brings them closer to face-to-face communication, and they believe that these moving images satisfy needs that text messages can't, such as expressing feelings or displaying metacommunicative signals.

2. **Searchers** The main motivation for searchers is their interest in the technical possibilities of GIFs and the humor of GIFs. They enjoy the process of finding the perfect GIF, although more for personal entertainment than for connecting with others. They feel uncomfortable using GIFs in romantic situations. For them, GIFs are useful only for their playful tone, not for serious topics. As searchers, they enjoy the challenge, but prefer GIFs that are quick and easy to share. GIF use among this group is highly situation-dependent and is mainly associated with humorous contexts.

3. **Referentialists:** The main motivations for referentialists are technical possibilities, cultural connection, and the humor of GIFs. For them, GIFs were used as an easily accessible way to reference popular culture. Although

they did not like the process of searching for GIFs, they often used them to connect with friends and were particularly interested in the pop culture of movies and television shows. They knew all the characters in these shows and felt that GIFs were funnier because they were related to these TV shows. For referentialist users, finding the perfect GIF was similar to avoiding awkward situations, as not having one could create socially awkward situations for them.

## Research

The focus of the research was to identify user groups related to GIFs, determine the characteristics that separate these groups from each other, and identify any sub-categories that may exist. The study focused on getting a deeper insight into the behavior, attitudes, and motivations of GIF users, with a special focus on the enthusiast, searcher, and referentialist categories mentioned by Church, King, and their colleagues in their study *“Relating, searching, and referencing: Assessing the appeal of using GIFs to Communicate.”*

From a methodological perspective, I used a survey in which I took and used statements belonging to different categories. I took these statements from the research mentioned above, based on what statements were associated with the different categories. The preliminary research assigned 12 statements to each group, which could be either positive or negative. These statements focused on different aspects, such as the motivations for using GIFs, such as emotional expression, search mechanism, or even humor.

The purpose of the survey was to assess participants attitudes toward GIFs and to enable me to identify and understand users belonging to different categories, as well as the visible differences between groups. In the survey, I examined participants' attitudes on a two-point scale of positive and negative attitudes. According to the instructions, participants had to answer “yes” if the statement described them and “no” if it did not.

Analyzing the proportion of positive and negative responses helped me to group and differentiate the attitudes and preferences of different user groups.

For the research, I selected the sample using expert sampling. The individuals in the sample had to meet several criteria. First and most importantly, they belonged to the same age group, were university students

aged 19-21, and were studying communication. There was no difference between the genders, both women and men could be included in the sample. In addition, it was important that the participants had prior knowledge of visual emoticons, with particular regard to GIFs, learned at university, and were familiar with their role in identity, their communication functions, etc. In addition, I also put together a negative group with similar characteristics to the main sample, but without previous knowledge of GIFs. I also selected 19-21-year-old university students from the communication department for the negative group, and there was no difference in the gender distribution. The purpose of the negative group was to allow the research to compare the reactions and attitudes of the two groups towards GIFs and to assess the impact of prior knowledge on the phenomena under investigation.

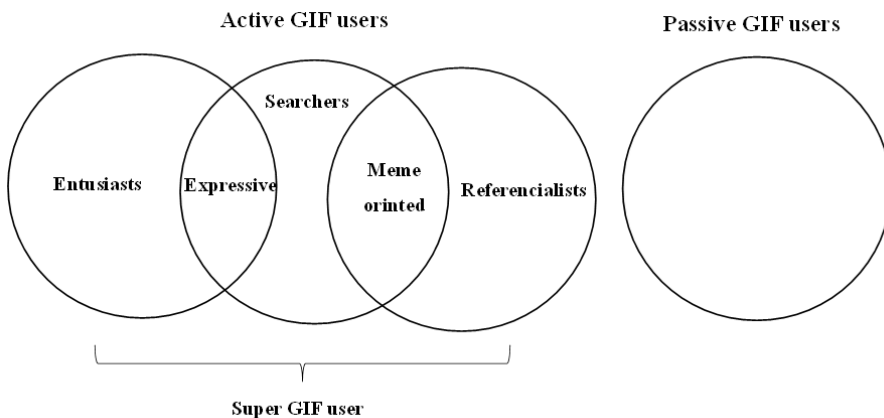
A total of 80 participants completed the survey, of whom two-thirds had prior knowledge of GIFs, while one-third did not. This information proved particularly useful as it allowed the participants to be divided into groups, which could then be compared during the analysis. The survey was not anonymous, as I will use the categorization to select interviewees for the next phase of my larger research. The lack of anonymity allowed me to identify and compare these groups.

During the analysis, I analyzed in detail the yes/no responses to the statements assigned to the GIF user categories defined by Church. For each statement with a positive value, I added a point for a yes response and ignored a no response. I used the same method for statements with a negative value. I added up these points for each participant to see how close they were to each category. I took different criteria into account when analyzing the results. The maximum possible score for each category was 12, and I only considered scores above 7 in my analysis. This helped me to exclude lower scores that may be less relevant to the analysis and to consider only outstanding or strong characteristics. If there were similar scores across multiple categories, but one category was at least 3 points higher than the other, I considered that category to be dominant, while I considered the lower value to be secondary. This approach allowed me to highlight the categories that were most characteristic or decisive for the participants.

## Analysis

During the analysis, I first identified two main categories: active and passive GIF users. I considered active users to be those who met the criteria I had set, that is, those with a value higher than 7, while I considered passive users to be those who did not meet the specified criteria. At this stage of the research, I did not find any significant differences between the different groups participating in the sample. During the analysis, I found identical categories in both groups. However, in the course of further research, the most significant difference in the sample will be explored, namely that one group has prior theoretical knowledge of visual emoticons, especially GIFs, and is aware of their role in identity, their communicative functions, etc.

However, the analysis required a more refined resolution, so I took into account the enthusiastic, referentialist, and searching categories known from Church's research. Of these, the enthusiastic category was predominantly present in the analysis, followed by the searching category, and finally, to a small extent, the referentialist group.



**Figure 1.** GIF user categories (based on research results – Own editing)

Based on the results, in addition to the pure categories, some combinations were also observed. There were few cases where the values of one category were dominant, it was much more common for two or possibly all three values to exceed the specified criterion of 7, or for there to be no greater difference between the values than 3.

### 1. Expressive

One such category mix was between the enthusiastic and searching categories. I identified these cases as the expressive category and defined them as individuals who use GIFs not only as a digital accessory but also as an important element of communication, which allows them to show their mood, reaction, and style. The defining statements of this category show that for users, the use of GIFs is not a functional but an expressive act that serves to represent themselves and display emotional intensity. Positive statements from the enthusiastic category mainly focus on the role of emotional expression and connection. Among the statements is the following *“I like to use GIFs to show my 'reaction when...' in hypothetical situations.”* describes the performative nature of expressive GIF use. The user represents their own personal responses, feelings, or attitudes through visual reactions, thereby constructing an emotional narrative. This statement also appears among the statements in the search category, but here it is more of a tool for connecting visual flexibility and personal creativity. In addition to expressing emotions, there is also a partner-oriented nature, which is mainly used to communicate humour, this is reflected in the statement *“I love finding the perfect GIF to make my friends laugh,”* which appeared at a high rate in the expressive category, showing that humor and laughter strengthen bonds. The third statement from the enthusiastic category that achieved a high percentage is *“I like GIFs because they can convey much more emotion than emojis,”* which shows that those in the expressive category are aware that there are still visual expressions of emotion in the online environment, yet they consider GIFs to be more expressive. The statements in the enthusiastic category are complemented by positive statements from the searcher type, which achieved high scores. One of these statements is *“I like that GIFs are fast and don't take too much time to view,”* which refers to the speed and ease of use of digital communication, as GIFs are instant, short, and easily interpretable visual units.

In profiling the expressive category, the values of negative statements also help, pointing to the conscious use of GIFs. Negative statements should be interpreted in reverse, as in these cases the “no” answer was the correct one. In this reverse interpretation, users see not limitations but possibilities in GIFs. Statements such as *“I don't think GIFs are useful for expressing my emotions in text conversations”* indicate that the expressive category attributes

emotional value to GIFs and sees potential in GIFs to respond more playfully and quickly in emotional situations. GIFs thus become not a sign of a lack of authenticity, but a form of digital authenticity, that is, a natural means of expressing emotions in the online space. In this case, the statement *“I don’t like GIFs because they are part of internet culture”* indicates that this group is not bothered by the fact that GIFs are part of mainstream culture. The user does not reject the social codes of internet culture, but identifies with them and values GIFs as a cultural phenomenon. This attitude reflects a sense of belonging to digital communities and an embrace of collective identity. Similarly, *“I feel like I have to use GIFs because everyone else is using them”* also refers to belonging to a community or the opportunity to connect. Thus, GIF use becomes a natural part of social interactions, in which the individual does not lose themselves but expresses themselves through shared codes.

## **2. Meme oriented**

Another interesting combination that appeared during the analysis is between the searcher and referentialist categories, which I have named the “meme oriented” category. Its most important characteristic is that this category does not use GIFs to express emotions, but rather as a cultural and intertextual communication tool. The name of this category was inspired by memes, as GIFs are interpreted as carriers of shared knowledge, pop culture references, and irony. For them, GIFs are not an expressive medium in themselves, but a conveyor of layers of meaning, in which familiar references to films, TV series, or memes form the basis of collective understanding.

One of the key characteristics of meme GIF users, which can be identified from the positive statements in the results, is that this category values the technological and visual effectiveness of the GIF format. The popularity of the statement *“I like that GIFs are fast and don’t take too long to view”* shows that for meme creators, the brevity and dynamism of GIFs make them suitable for use as instant reactions, similar to the quick spread of memes. Another search term that appears is *“I like how GIFs act out what I want to say,”* which emphasizes the expressive function of GIFs, that is, GIFs help users express their gestures or reactions. This includes statements from the reference category, such as *“I love GIFs for their pop culture references,”* highlighting that users choose GIFs not only for their visual effectiveness, but also for their familiarity and shared coding. Pop culture references from

GIFs become social reference points that create a kind of an “inside joke.” Closely related to this is *“I use GIFs as inside jokes between myself and my friends,”* which, on the one hand, points to this shared knowledge and, on the other hand, highlights that it becomes a cultural language for the group of meme GIF users. This is reinforced by the statement *“I mainly use GIFs from movies or TV shows that my friends and I are fans of,”* which reflects shared fandom and the visual expression of online community identity.

Negative statements also point to important characteristics of the category. The statement *“I don't think GIFs are useful for expressing my emotions in text-based conversations”* points to the fact that meme creators also see potential in GIFs as expressions of emotion in the online environment. Another statement, the reverse of *“Using GIFs is an easy way out for me when I don't know what else to say,”* shows that this category is not used to maintain conversation, so they are characterized by a kind of isolation.

When analyzing the mixed categories, interesting observations were made about how certain categories mix or do not mix with each other. Analysis of the data revealed that there is no significant mixing between the enthusiastic and referentialist categories. This may be explained by the fact that these two categories have extremely different characteristics in terms of factors such as motivation, communication functions, and cultural connections.

For enthusiastic GIFs are used because of their strong emotional and communicative potential, helping to make their nonverbal cues in conversation similar to those in face-to-face situations. On the other hand, the main characteristic of the referentialist group is that they pay special attention to background knowledge and use GIFs primarily for the easily accessible display of references to popular culture. This difference may explain why there is no significant overlap between the two categories. Enthusiastic and referentialists use GIFs for different purposes and motivations, so the differences between the two groups outweigh the possibility of mixing.

### *Super GIF User*

Although not all categories were mixed during double mixing, the analysis revealed that there were participants who scored high in all three categories, meaning they were enthusiastic, searchers, and referentialists at the same time. I considered them to be the “super GIF user” category, users who consciously use GIFs and have advanced digital literacy and digital

competencies. For this category, GIFs are not just an accessory, but rather a multimodal means of expression that serves to express emotional identification, humorous interaction, and pop culture references all at once. At the intersection of the three categories, super GIF users have multi-level competence in digital communication: they are able to read and recreate visual codes while understanding their cultural, technological, and social dimensions.

According to the positive statements from the enthusiastic category, GIFs are primarily a means of expressing emotions for super GIF users, as reflected in the statement, *“I love GIFs because they can convey much more emotion than emojis.”* Based on this, GIFs are not just simple illustrations, but conveyors of emotional nuance and expressiveness. Similarly, the statement *“I love that GIFs act out what I want to say”* emphasizes the performative role of GIFs, according to which GIFs are not only present as reactions, but are also capable of creating visual narratives by acting on behalf of the user. The statement *“I love finding the perfect GIF to make my friends laugh”* stands out among the positive statements in the enthusiastic category, highlighting both humor and social connection, in which GIFs become an important element of shared laughter or community building.

The statements in the search category mainly emphasize digital skills, with the super GIF user category seeing potential in GIFs, highlighting one of their main features, looped motion. This is demonstrated by the statement *“I like GIFs because of their animation and looped movement,”* which evaluates GIFs from an aesthetic point of view. The continuum of movement not only maintains attention but also creates a visual experience. In addition, the statement *“I like that GIFs are fast and don't take too long to view”* also highlights the technical feature of GIFs, namely that they can be consumed quickly and easily.

The positive statements in the third, referential category point to characteristics of super GIF users such as background knowledge or social dimensions. The statement *“I like GIFs because of their pop culture references”* points to the pop culture literacy behind GIFs, which links the message to a particular film, series, or community experience. Related to this is the fact that super GIF users refer to pop culture knowledge and the internal codes of the community, where GIFs are a visual language of understanding and belonging, as indicated by the statement *“I use GIFs as inside jokes between myself and my friends.”* The statement *“GIFs from movies and TV shows that I love*

*make me feel nostalgic*” reflects digital nostalgia and an emotional attachment to moving images, which can be considered an important tool for community building.

Furthermore, the negative statements of the enthusiastic, searching, and referential user categories also draw attention to important characteristics of super GIF users. First, based on the reverse interpretation of the statement “*GIFs don't grab my attention as much as emojis*” from the enthusiastic and referential categories, super GIF users view GIFs as a more complex visual code that can convey more layers of meaning than emojis.

Similarly, in the statement “*I don't like GIFs because they are part of internet culture,*” which also comes from the enthusiastic and referentialist categories, super GIF users don't separate themselves, but identify with the common digital culture and see GIFs as a symbol of cultural participation and creative community. Closely related to this is the negative statement from the enthusiastic category, “*I feel like I have to use GIFs because everyone else is using them,*” which, when interpreted in reverse, expresses the intention to connect. In this interpretation, the super GIF user appears not as a participant in popular culture, but as an active shaper of digital community language. The shared use of GIFs is not a compulsion, but an opportunity for interaction in which the user participates creatively and consciously.

The negative statement in the search category, “*I use GIFs because videos sometimes take too long to load,*” highlights the functionality of GIFs. In this context, GIFs are not a “weaker” alternative to video, but a fast, condensed, and visually effective means of communication. Super GIF users consciously use the temporal and formal characteristics of the medium to reinforce their message. The negative statement “*I use GIFs to feel closer to people when I send them messages,*” which comes from the searcher and referentialist categories, points to the fact that for super GIF users, GIFs are not only an aesthetic medium, but also an interpersonal one. Thus, it can be seen as a communicative gesture that combines visual humor, irony, and emotion, all of which play a role in making connections in the digital space. Finally, the statement “*I don't think GIFs are useful for expressing my emotions in text conversations*” comes from the negative statements in the search category, which, in reverse interpretation, defines GIFs as a language that focuses on the expression of emotion. Based on this, the super GIF user sees GIFs as an opportunity in the field of digital emotional expression and is aware that communication in the online space needs metalinguistic cues, and GIFs can be suitable for this.

During my analysis, I found that humor is a prominent factor in all categories, whether purely enthusiastic, referential, or mixed. In every category, it is important that users use GIFs as a source of humor. This can be explained by the fact that humor plays a general and decisive role in communication, regardless of the category to which the given form of communication or person belongs. In the pure enthusiastic and referentialist categories, humor is also an important part of communication. In the enthusiastic category, for example, humor can help strengthen relationships and express identity, while in the referential category, humor generally plays a role in demonstrating pop culture literacy, thus making the use of GIFs in a humorous context particularly important within these categories. In the case of mixed categories, it is also important to note that humor is a key factor. Even though these categories generally combine purely enthusiastic and referentialist characteristics, humor continues to play a prominent role in interpersonal relationships and message transmission.

## **Conclusions**

The research presented a comprehensive overview of how GIFs are not only complex in themselves, but also interpreted in a wide variety of ways by users. By further developing the initial research, I was able to identify new, more nuanced user categories that carry additional shades and dimensions beyond the extremes. It can be concluded that pure categories appeared less frequently than mixed categories, which can be partly attributed to the fact that the enthusiastic and referentialist categories, in particular, can be considered extreme. The enthusiastic category includes those who place particular emphasis on partner-oriented communication and, to this end, enrich it with nonverbal cues and the presentation of their unique identity. On the other hand, the referentialist category is more self-serving, as the use of GIFs serves not to identify with others, but rather to showcase pop culture literacy. The searcher category is in the middle and is not exclusively about motivation and communication goals, but rather about the use of technological competencies. Based on the creation of subcategories, it can be said that GIF users are more complex and that GIFs as a visual tool can help to convey nonverbal signs, as seen in expressive GIF users, or can help to belong to a community, as observed in the meme-oriented category. What complements the underlying research is the mixture of clear categories,

that is, enthusiastic, searching, and referential categories, which I have named super GIF users. This category contains the important characteristics of all three categories, and it can be said that this is the category that exploits the full potential of GIFs and has all the skills necessary for using GIFs, both in terms of usage and understanding.

Based on the results of my research, I have come to the conclusion that GIF user categories form a complex system that shows similarities to users of other media, such as television, radio, or internet user communities. This approach allows for a deeper understanding and promotes further research in this area.

## References

- Church, S. H., King, J., Robinson, T., & Callahan, C. (2023). Relating, searching, and referencing: Assessing the appeal of using GIFs to communicate. *Convergence*, 29(3), 730–745. <https://doi.org/10.1177/1354856523115507>
- Coen, J. (2016, August 10). *History of the GIF*. Mashable. <http://mashable.com/2016/08/10/history-of-the-gif/>
- Eppink, J. (2014). A brief history of the GIF (so far). *Journal of Visual Culture*, 13(3), 298–306. <https://doi.org/10.1177/1470412914553365>
- Highfield, T., & Milner, R. M. (2017). Never gonna GIF you up: Analyzing the cultural significance of the animated GIF. *Social Media + Society*, 3(3). <https://doi.org/10.1177/2056305117725223>
- Meier, M. (2018, February 1). *Instagram supports Stories GIF stickers, WhatsApp integration: Important Instagram updates of January every marketer should know*. Medium. <https://medium.com/instaheads-blog/important-instagram-updates-of-january-every-marketer-should-know-fe13bc5166d7>
- Veszelszki, Á. (2015a). A netnyelvészet terminológiai és kronológiai áttekintése. *Magyar Nyelvőr*, 139(1), 23–42.
- Veszelszki, Á. (2015b). Emoticons vs. reaction-gifs: Non-verbal communication on the internet from the aspects of visibility, verbiage and time. In A. Benedek & K. Nyíri (Eds.), *Beyond words: Pictures, parables, paradoxes* (Visual Learning, Vol. 5, pp. 131–145). Peter Lang.
- Wagner, A. (2021). The postdigital emergence of memes and GIFs: Meaning, discourse, and hypernarrative creativity. *Postdigital Science and Education*, 3, 831–850. <https://doi.org/10.1007/s42438-020-00160-1>
- Wang, M., Bhuiyan, M. E., Rho, E. H. R., Lee, S. W., et al. (2024). Understanding the relationship between social identity and self-expression through animated GIFs on social media. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1), 1–30. <https://doi.org/10.1145/3641031>

## Charting Ethical Communication in Relation to Integrity in Scientific Research

Georgiana Silvia LEOTESCU<sup>1</sup> 

**Abstract.** The current study endeavours to highlight the importance of being aware of and adhering to regulations regarding ethics and integrity while conducting scientific research. It underlines that complying with ethical norms and pursuing strong moral values has become a prerequisite to pristine professional conduct, a sign of respect towards the academic community. As such, the paper firstly offers a general overview concerning key factors that should be taken into consideration so that deviant, unethical decisions do not reflect poorly on any profession. There is also a section dedicated to the impact of research misconduct and case studies are presented in order to address the breaches that can appear. Accordingly, it dwells on multiple international guidelines that have been published, while also reflecting on how our own country has contributed to creating reference frameworks that promote a responsible direction in scientific research. Starting with the European Charter for Researchers issued in 2005 by the European Commission, the paper underscores major guidelines of ethically justified scientific practice in countries such as Belgium, the Netherlands, or the UK. It lastly focuses on ethical considerations in Romania, especially in relation to current implications regarding the use of artificial intelligence, pondering on how recent technological developments have impacted humans' adherence to ethical principles and values.

**Keywords:** ethics, integrity, scientific research, guidelines, AI regulation.

---

<sup>1</sup> Lecturer, Department of Education and Communication Sciences, University of Craiova, Romania, georgiana.leotescu@edu.ucv.ro



## Introduction

Ethical considerations should be at the core of every researcher's scientific endeavor to ensure that (s)he adequately fulfils the mission of making a contribution to the research field(s) that (s)he is interested in. One should bear in mind that taking into consideration all the ethical aspects does not only reinforce the visibility and recognition of the conducted research, but it also confirms the level of respect towards the academic community, as a whole, and the determination not to lower one's standards, at an individual level. In fact, being honest and having strong moral principles that you refuse to change, no matter the situation or context that you are confronted with, stands for unimpeachable integrity and character.

In scientific research, integrity is guaranteed by adhering to values that both researchers and research organizations must meet through principles and / or guidelines. According to a document issued by the European Commission in 2018, which aims to signal the importance of *Ethics in Social Sciences and Humanities*, two different ethical dimensions of research can be observed: “«procedural ethics», pertaining to the aspects of compliance in performing research, and «ethics in practice», the everyday ethical issues that arise while doing research.” (2018, p. 3). As such, it is of utmost importance not only to consider the impact of acknowledging and complying with recognized principles but also to achieve the desired results by actually conducting the research in a responsible manner.

Responsibility is closely linked to academic freedom, that is, no matter if researchers can choose from a wide array of topics and set their aims independently, they must reach their goals conscientiously. *The European Charter for Researchers* clarifies this aspect, albeit referring to research freedom instead of academic liberty as a general principle to be applied: “Researchers should focus their research for the good of mankind and for expanding the frontiers of scientific knowledge, while enjoying the freedom of thought and expression, and the freedom to identify methods by which problems are solved, according to recognised ethical principles and practices. Researchers should, however, recognise the limitations to this freedom that could arise as a result of particular research circumstances (including supervision / guidance / management) or operational constraints, e.g. for budgetary or infrastructural reasons or, especially in the industrial sector, for reasons of intellectual property protection. Such limitations should not, however, contravene recognised ethical principles and practices, to which researchers have to adhere.” (2005, p. 11).

## Literature Review

Ethical concerns have been addressed in various fields and across multiple areas worldwide. In Romania, the Ministry of Education and Research requested the National Council of Ethics in Scientific Research, Technological Development and Innovation to issue a document that would contribute to the consolidation of responsible science, values, and good practices of integrity within the scientific research communities that are part of the research and development system in Romania. The national regulations (*Ghid de integritate în cercetarea științifică*) were published in 2020 and feature four values that are pivotal in scientific research: honesty, responsibility, verifiability and validity (2020, pp. 5-6).

In terms of international guidelines, however, there are several documents which stipulate the correct decisions that researchers have to make regardless of the stage of research process in which (s)he finds himself / herself, thus, showing respect for the entire academic community. In October 2018, The European Commission issued the guidelines entitled *Ethics in Social Science and Humanities*, focusing much of the intended content methods of mitigating ethical problems that could appear especially when social sciences and humanities methodology is employed. As such, the document proposes the following central principles correlated especially with EU-funded research: “respecting human dignity and integrity, ensuring honesty and transparency towards research subjects, respecting individual autonomy and obtaining free and informed consent (as well as assent whenever relevant), protecting vulnerable individuals, ensuring privacy and confidentiality, promoting justice and inclusiveness, minimising harm and maximising benefit, sharing the benefits with disadvantaged populations, especially if the research is being carried out in developing countries, respecting and protecting the environment and future generations.” (2018, p. 5).

The document also provides useful insight into ethical dimensions of research methodology, informed consent, findings outside the scope of research (categorized as “unintended / unexpected / incidental findings”, data protection and privacy, sites of research (with a special focus on risks, harm and risk assessment), misuse of research, ethics approval in social sciences and humanities research, management of ethics issues and others as well.

This is actually a reference framework applied specifically to social sciences and humanities and reinforces the general principles and requirements enumerated in the *European Charter for Researchers* issued by the European

Commission on the 11<sup>th</sup> March 2005. Addressing researchers, employers and funders alike, this framework proposed the following general principles and requirements applicable to researchers: research freedom, ethical principles, professional responsibility, professional attitude, contractual and legal obligations, accountability, good practice in research, dissemination, exploitation of results, public engagement, relation with supervisors, supervision and managerial duties and continuing professional development. Concurrently, employers and funders must adhere to another category of general principles and requirements such as: recognition of the profession, non-discrimination, research environment, working conditions, stability and permanence of employment, funding and salaries, gender balance, career development, value of mobility, access to research training and continuous development, access to career advice, intellectual property rights, co-authorship, supervision, teaching, evaluation/appraisal systems, complaints/appeals, participation in decision-making bodies and recruitment.

There are also reference frameworks which establish major guidelines of ethically justified scientific practice in certain countries. For example, the *Code of Ethics for Scientific Research in Belgium* (2009) comprises six values which are presented as three pairs. The first one refers to rigour and caution and pinpoints the obligation to avoid generating any negative effects. The second one explains that reliability and verifiability are at the core of conducting trustworthy research that can be reproduced, if necessary. The third pair signals the importance of independence and impartiality, highlighting that rules of a scientific nature guide researchers who must rely on objectivity throughout the entire research process.

In 2018, a new document is published as an updated version of the *Netherlands Code of Conduct for Academic Practice* (2004). As such, the *Netherlands Code of Conduct for Research Integrity* signals that more attention should be devoted to the importance of research integrity especially in the context of modern knowledge societies and the continuous development that we are experiencing. The main principles that are described in this document feature: honesty, scrupulousness, transparency, independence and responsibility. We can observe that two of them, namely honesty and responsibility also feature in our national set of guidelines published two years after, in 2020.

Prior to both reference frameworks, the *Danish Code of Conduct for Research Integrity* (2014) sets only three principles of research integrity which should be applied in all the stages of research: honesty, transparency and accountability. Once more, honesty appears to be at the forefront of the

discussions in connection to high-quality research and education, overall. Nevertheless, the Danish Code of Conduct supplements the three principles with six basic standards while conducting responsible research: research planning and conduct, data management, publication and communication, authorship, collaborative research and conflicts of interest.

Recently, the UK Research Integrity Office published the *Code of Practice for Research* with the intention of complementing the existing guidance on research conduct, which also features *The Concordat to Support Research Integrity* with its Five Commitments (2019 edition): “maintaining the highest standards”, “ethical, legal and other frameworks”, “research culture”, “dealing with research misconduct” and “strengthening research integrity”. The *Code of Practice for Research* includes, first of all, a checklist for researchers which is, basically, a set of questions that they must provide answers for but only after they have explored the guidance available in another section of the document. In addition to the five commitments that we have already mentioned, the revised edition pinpoints other key elements which foster a healthy research culture such as: research vigour, reflection, transparency and environmental impact. (2023, p. 7). The last part of the document is dedicated to highlighting standards for organisations and researchers alike.

### **Ethical misconduct in research. Impact and case studies**

There are multiple implications if research misconduct is investigated and proven, since we all know that lying or cheating can have negative consequences no matter the researchers’ motivation(s) behind this improper behaviour. Accordingly, there can be various penalties for scientific misconduct but only if one discovers the faulty behaviour and provides solid evidence in that specific misconduct case. For example, in some researchers’ view, “fabrication and falsification contaminate scientific literature with incorrect information, which wastes funds, hampers medical and technological development, and poses risks to patients and consumers, while plagiarism deprives authors of credit for their work” (DuBois et al., 2013, p. 321). It is also important to mention that “misconduct cases are typically shrouded in secrecy, and studies on professional wrongdoing suggest that the causes of misconduct are complex.” (DuBois et al., *idem*). Accordingly, if incorrect (unethical) actions are undertaken by researchers, there is a risk that “in some cases, the extent of the penalty is that the researcher’s reputation is forever tainted” (D’Angelo, 2018, p. 35).

The revised edition of *The European Code of Conduct for Research Integrity* also appeared in 2023 with the following fundamental principles to guide not only individuals, but also institutions and organisations in their quest to ensure research integrity: reliability, honesty, respect and accountability. Once again, honesty appears to be a crucial factor in promoting an ethical mindset. Good research practices are also described in these guidelines in connection to eight different contexts: “research environment”, “training, supervision and mentoring”, “research procedures”, “safeguards”, “data practices and management”, “collaborative working”, “publication, dissemination and authorship” and “reviewing and assessment”. (2023, p. 6). The last section of the reference framework signals violations of research integrity as forms of “research misconduct”, such as: fabrication, falsification, plagiarism, failure to meet legal, ethical and professional obligations, misrepresentation of [data, involvement, interests, qualifications, publication history], improper dealing with allegations of misconduct.

In the context of misconduct, Marijke Van Buggenhout, Jenneke Christiaens and Serge Gutwirth submitted a document in 2017 entitled *Promoting Integrity as an Integral Dimension of Excellence in Research. Final Report on the Incidence of Misconduct*. The document was resubmitted after the final review report with the abbreviated title PRINTEGER and besides the background information on the measurement of misconduct in science, it also features reports on research misconduct in the following countries: The Netherlands, Estonia, Norway, Belgium, Italy and Great Britain. The data collection method and the results are also provided in this comprehensive study. Additionally, in 2019 the ENRIO Handbook was published, providing another set of as *Recommendations for the Investigation of Research Misconduct*. In the Preamble, we can discover that ENRIO stands for the European Network of Research Integrity Offices, which was founded as an initiative by UKRIO (the UK Research Integrity Office). The discrepancies between the European Code of Conduct (Coc) and this handbook are also emphasized according to the detailed section which is added to the latter, comprising “detailed practical recommendations on how to deal with research misconduct and other unacceptable practices”. (2019, p. i).

In *Ethics in Science: Ethical Misconduct in Scientific Research* (2018), D’Angelo explores each type of misconduct as an effective strategy to prevent such cases of ethical violation. Since the main focus of this paper is to provide an overview of ethics and integrity and research, we will not concentrate on

exploring each type of misconduct in detail, but present the historical cases that he enumerates. Thus, in Chapter 9, he starts with “intentional negligence in acknowledgment of previous work done” and continues to provide explanations for “deliberate fabrication of data”, “deliberate omission of known data that doesn’t agree with hypotheses”, “passing another researcher’s data as one’s own”, “publication of results without consent of all the researchers”, “failure to acknowledge all the researchers who performed the work”, “conflict of interests”, “repeated publication of too similar results”, “breach of confidentiality” and “misrepresenting others work” (D’Angelo, 2018, p. 101-106). In the following pages, he also covers some case studies applying the following methodology for each case under discussion: summary, starting point, resolution, questions to ponder and sources. The inclusion of questions to ponder is explained in the first paragraph of the chapter: “Unlike the examples covered previously, for which it was clear that misconduct occurred, and penalties were levied, some of the cases that follow in this section are much less clear-cut. It is the intent behind this presentation that these cases stimulate conversation (and perhaps even lively debate) about whether misconduct has taken place (D’Angelo, 2018, p. 109). The fifteen analysed cases are the following: Darwin and Wallace, Rangaswamy Srinivasan – Visx Patent Dispute, Schwartz and Mirkin, Corey and Woodward, Córdova, Woodward and Quinine, DNA, David Baltimore and Theresza Imanishi-Kari, John Fenn-Yale Patent Dispute, Vioxx, Autism and Measels, Mumps, and Rubella Vaccine, Italian Earthquake Scientists, The Red River Floods, Animals Lost During Storms or Other Acts of Nature and Transfer of a Patent to an Indian Tribe to Avoid its Invalidation. Another significant article that we find significant to mention at this point, especially due to the authors’ historiometric approach (statistical analysis of coded data from historical narratives) is “Understanding Research Misconduct: A Comparative Analysis of 120 Cases of Professional Wrongdoing” published in 2013.

Our research also signals another valuable reference in terms of addressing breaches of research and surveying publication ethics. Founded in 1997, as “a voluntary body providing a discussion forum and advice for scientific editors”, COPE (Committee on Publication Ethics) “aims to find practical ways of dealing with the issues, and to develop good practice” (COPE, *Guidelines on Good Publication practice*, p. 43). As such, the *Guidelines on Good Publication practice* “should be useful for authors, editors, editorial board members, readers, owners of journals, and publishers. Intellectual honesty

should be actively encouraged in all medical and scientific courses of study, and used to inform publication ethics and prevent misconduct.” (COPE, *idem*). There are ten guidelines proposed to ensure the ethics of scientific publishing (study design and ethical approval, data analysis, authorship, conflict of interests, the peerreview process, redundant publication, plagiarism, duties of editors, media relations, advertising) and a special section dedicated to dealing with misconduct. For the purpose of our research, we also wish to highlight that this forum has got a special section where specific publication ethics issues are presented and discussed by COPE Members. The particular cases under consideration are anonymised and advice is given during COPE Forum meetings, from the COPE Council and COPE Members across various regions and disciplines. It is important to point out that the case studies we have chosen have got an on-going status, which means that the final conclusions have not yet been drawn.

## **Current Trends and Developments**

Scientific research integrity in Romania is regulated by Law no. 206 / 2004 on proper conduct in scientific research. The key legal aspects and definitions of misconduct are outlined in this document, and the law identifies three main categories of violations, namely plagiarism, data fabrication, and data falsification. The legislation aims to provide ethical standards and encourage responsible behaviour among researchers, including universities and research institutes. There are multiple practices that are considered inconsistent with appropriate conduct in research and development, such as: concealing or omitting unfavourable or undesired research results, fabricating research data, substituting actual findings with fabricated or fictitious data, intentionally misinterpreting research results or drawing misleading conclusions, plagiarizing the findings or publications of other researchers, deliberately misrepresenting or distorting the research outcomes of others, incorrect or unethical attribution of authorship, inclusion of false or misleading information in grant or research funding applications, failure to disclose actual or potential conflicts of interest, misappropriation or unauthorized use of research funds, failure to register or report research results when required, lack of objectivity in the evaluation process and failure to comply with confidentiality obligations, or

repeated publication of the same research findings while presenting them as novel scientific contributions.

In cases where breaches of good conduct in research and development are identified and proven, the National Ethics Council may recommend the application of one or more of the following sanctions: removal of the individual or individuals from the research project implementation team, replacement of the project manager, retraction and /or correction of all publications produced in violation of the principles of good research conduct, issuance of a formal written warning, demotion to a lower position or rank, temporary suspension from professional duties, dismissal from employment.

In order to complement this legislation, the Ministry of Education and Research requested the National Council of Ethics in Scientific Research, Technological Development and Innovation (CNECSDTI – Consiliul Național de Etică a Cercetării Științifice, Dezvoltării Tehnologice și Inovării) to issue a document that would contribute to the consolidation of responsible science, values, and good practices of integrity within the scientific research communities that are part of the research and development system in Romania. The National Council of Ethics in Scientific Research, Technological Development and Innovation (CNECSDTI) of Romania is an advisory body without legal personality operating under the state authority responsible for research and development, as provided in Articles 5-8 of Law No. 206 / 2004. The National Authority provides administrative support to the CNECSDTI Research Department. As such, the Integrity Guide (*Ghid de integritate în cercetarea științifică*) was published on 12 November 2020. The national regulations feature four values that are pivotal in scientific research: honesty, verifiability, responsibility and validity (2020, pp. 5-6).

According to this guide, honesty in research requires adherence to the following principles: the rejection and prevention of all forms of intellectual misconduct (including plagiarism), the rejection and prevention of any practice involving the fabrication of data or the presentation of fictitious information as authentic research findings, the rejection and prevention of any form of data falsification, alteration, or manipulation, the explicit acknowledgment of each author's contributions to the development and creation of a collaborative scientific work, the clear identification of the specific contributions made by individual authors, where applicable, within the content of the publication, the proper recognition and citation of the author's previously published works to prevent self-plagiarism (2020, pp. 6-7). The honesty principle also

refers to the complete, accurate, and transparent reporting of research findings, irrespective of whether the results (1) support the initial research hypotheses or (2) align with the interests of research sponsors, funding bodies, or the researchers themselves. Last but not least, it involves respect for the public as the ultimate beneficiary and end user of research outcomes, which requires avoiding any form of manipulation through inadequate research design, inappropriate methodological choices, or the selective presentation of findings that may be influenced by the interests of funding organizations.

The second value that should be respected in the research process, verifiability, requires that research reports and the methodological sections of scientific publications provide a comprehensive description of the procedures used for data collection or generation, including information on the location, timeframe, and methods employed. In the case of qualitative social research, this requirement also includes providing relevant descriptive information regarding the socio-demographic characteristics of the interviewed participants. Verifiability also accounts for the presentation and dissemination of research findings, which must strictly adhere to confidentiality and data protection principles concerning the individuals or groups involved in the study. Also, any modifications made to datasets during the analytical process, such as the removal of outliers, the replacement of missing values for specific variables, or other forms of data transformation, must be clearly reported, explained, and scientifically justified (2020, p. 7). Also, the reporting of quantitative analysis results should include a clear specification of any data weighting procedures or techniques used for smoothing frequency distributions of particular variables while research data, whether quantitative or qualitative, should be archived according to standardized procedures and retained for defined periods of time.

The responsibility principle implies obtaining the informed consent of research participants before their involvement in data provision or experimental procedures constitutes an essential element of the aforementioned accountability practices, the establishment of institutional mechanisms for the prevention and sanctioning of deviations from the standards of good research practice within the relevant field, the selection of research topics addressing issues of significant social and human relevance. Eventually, conducting scientific research in a responsible manner means communicating research findings that contribute to strengthening public trust in researchers and research institutions (2020, p. 8).

In order to ensure validity of the research process, the guidelines refer to the application of specific tests designed to assess the consistency of new research findings in relation to previous research results, established theoretical frameworks, or objective evaluation criteria. Also, valid research implies the use of multiple analytical methods applied to the same dataset, or the implementation of methodological variations, to evaluate the extent to which research conclusions remain stable and reliable under different analytical approaches (2020, p. 8).

The Higher Education Law No. 199/2023 establishes that compliance with the principles and standards of university ethics and academic deontology is a mandatory requirement for all educational, research, and administrative activities conducted within higher education institutions. In accordance with national legislation, each university is required to develop and implement its own Code of Ethics, which must be consistent with the rules and principles established by the National Framework Code of University Ethics and Deontology (*Codul-cadru de etică și deontologie universitară*).

Due to current trends and developments concerning the rise of generative AI, the Government of Romania established the Scientific and Ethical Council for Artificial Intelligence in May 2023. The Council functions under the coordination of the Ministry of Research, Innovation and Digitalization as an advisory body to support the ethical and responsible use of artificial intelligence (AI) technologies at the national level. The Council plays an advisory role which consists in providing scientific and ethical expertise to the Government and public authorities to support the appropriate implementation and governance of emerging technologies. Recognized experts in the field (such as Maria Luciana Axente, who has international experience in AI ethics and governance) are in charge of regulating and supervising the following key areas of action: supporting the development of a legislative framework that is responsive to technological innovation, promoting transparent, human-centered approaches to the development and deployment of AI systems, and addressing and mitigating the risks associated with the uncontrolled use of algorithmic technologies.

Prior to the initiative of establishing the aforementioned Council and in relation to the topic of artificial intelligence, it is worth signaling that the University Politehnica of Bucharest and the Research Institute for Artificial Intelligence of the Romanian Academy, among others, had already founded in April 2011 a nonprofit scientific association. Thus, ARIA (Asociația Română

pentru Inteligență Artificială) contributes to fostering the generation, development, and dissemination of knowledge and technological innovation in the domain of Artificial intelligence, with the mission to advance research and education in the field.

Our findings underscore that the *European Ethics Guidelines for Trustworthy AI* was the first and most comprehensive reference framework. It was presented in April 2019 and pinpointed the “lawful”, “ethical” and “robust” characteristics that trustworthy AI should possess.

Shortly after, in May, the OECD (Organisation for Economic Co-operation and Development) adopted the first intergovernmental standard on AI – the *Recommendation on Artificial Intelligence*. UNESCO also released its *Recommendation on the Ethics of Artificial Intelligence* which was adopted on the 23<sup>rd</sup> November 2021. Raising ethical concerns on the accelerating use of AI technologies, the UNESCO guidelines serve as “a standard-setting instrument developed through a global approach, based on international law, focusing on human dignity and human rights, as well as gender equality, social and economic justice and development, physical and mental well-being, diversity, interconnectedness, inclusiveness, and environmental and ecosystem protection can guide AI technologies in a responsible direction.” (2021, p. 5).

On the 30<sup>th</sup> November 2023, the Council of Europe released the “Guidelines on the responsible implementation of artificial intelligence systems in journalism”, which were adopted by the Steering Committee on Media and Information Society (CDMSI) on 30 November 2023. Hence, with the new policy framework, additional sectoral guidance is available, narrowing the earlier recommendations that were much global in scope. For example, the discrepancy between an artificial intelligence system and a journalistic artificial intelligence system is clearly stated, as the latter is “directly related to the business or practice of regularly producing information about contemporary affairs of public interest and importance, including the research and investigation tasks that underpin journalistic outputs”. Moreover, the definition underpins that “this can include, but is not limited to, large language models and generative AI when used for journalistic purposes and / or news organisations”, reinforcing, at the same time, that these systems “are not a single technology but a range of different, often interlinked, tools for automating specific tasks” (CDMSI, 2023, p. 3).

## Conclusion

As some scholars have already underlined, there is a fine line between good research practices and misconduct in scientific investigations. Unfortunately, there are also no universally accepted guidelines to distinguish unethical behaviour from actions which are ethically flawless (Wimmer & Dominik, 2014, p. 65). In fact, our research highlights several policy frameworks which have been developed to serve as recommendations or guidelines in order to inform the debate about ethics and the research process. Thus, an overview which we consider helpful encompasses the following regulations regarding ethics and integrity in research issued at both national and international levels, in a chronological order: *The European Charter for Researchers* (2005), *The Code of Ethics for Scientific Research in Belgium* (2009), *The Danish Code of Conduct for Research Integrity* (2014), *Ethics in Social Sciences and Humanities* (2018), *The Netherlands Code of Conduct for Academic Practice* (2018 - an updated version of the initial document published in 2004), *The Code of Practice for Research* (UKRIO, 2019), *Recommendations for the Investigation of Research Misconduct* (ENRIO, 2019), *Ghid the integritate în cercetarea științifică* (2020), *The European Code of Conduct For Research Integrity* (2023).

Recently, due to the fast pace at which technology has evolved, concerns regarding the ethical use of artificial intelligence have also been raised, thus resulting in policy frameworks and guidelines dedicated to AI regulation. Hence, by raising general awareness of the accelerating adoption of AI through the following sets of guidelines, we can ensure that ethical challenges are accounted for and discussed to avoid as much as possible shortcomings and inadequate deployment of AI systems that are incongruent with ethical principles: *The European Ethics Guidelines for Trustworthy AI* (2019), *The Recommendation on Artificial Intelligence* (OECD, 2019), *The Recommendation on the Ethics of Artificial Intelligence* (UNESCO, 2021), *Guidelines on the responsible implementation of artificial intelligence systems in journalism* (Council of Europe, 2023). We also signal the importance of establishing an advisory board in Romania in May 2023, namely the Scientific and Ethical Council for Artificial Intelligence (Consiliul Științific și de Etică în Inteligență Artificială), which could be interpreted as an act of alignment with international initiatives and commitment towards ethical and responsible deployment of AI.

## References

- Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique. (2009). *Codes of Ethics for Scientific Research in Belgium*.
- ALLEA. (2023). *The European Code of Conduct for Research Integrity*, Revised Edition 2023. Berlin. DOI 10.26356/ECOC. ISBN 978-3-9823562-3-5.
- Bouter, L. M., Tjldink, J., Axelsen, N., Martinson, B. C., & ter Riet, G. (2016). Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. *Research Integrity and Peer Review*, 1(1), 17.
- Committee on Publication Ethics (COPE). *Guidelines on Good Publication Practice*, available at <https://publicationethics.org/files/u7141/1999pdf13.pdf> (accessed on 17.01.2024).
- COPE (Committee on Publication Ethics), *Guidelines on Good Publication Practice*, the COPE Report 1999, pp. 43-47.
- Council of Europe. CDMSI. (2023). "Guidelines on the responsible implementation of artificial intelligence systems in journalism", adopted by the Steering Committee on Media and Information Societies (CDMSI) on 30 November 2023.
- D'Angelo, John. (2018). *Ethics in Science: Ethical Misconduct in Scientific Research*. Second Edition, Chapman and Hall/CRC.
- DuBois, J. M., Anderson, E. E., Chibnall, J., Carroll, K., Gibb, T., Ogbuka, C., Rubbelke, T. (2013). Understanding Research Misconduct: A Comparative Analysis of 120 Cases of Professional Wrongdoing, *Accountability in Research*, 20:5-6, 320-338, DOI: 10.1080/08989621.2013.822248.
- ENERI, European Commission. (2019). *Recommendations for the Investigation of Research Misconduct. ENRIO Handbook of Research Integrity*, Bonn; Vienna, ENERI; ENRIO.
- European Commission. (2018). *Ethics in Social Sciences and Humanities*.
- European Commission. (2005). *The European Charter for Researchers*, 1-23.
- Godecharle, S., Nemery, B., & Dierickx, K. (2014). Heterogeneity in European research integrity guidance: relying on values or norms? *Journal of Empirical Research on Human Research Ethics*, 9(3), 79-90.
- Ministerul Educației și Cercetării, Consiliul Național de Etică a Cercetării Științifice, Dezvoltării Tehnologice și Inovării. (2020). *Ghid de integritate în cercetarea științifică*.
- Ministry of Higher Education and Science. (2014). *The Danish Guide of Conduct for Research Integrity*.
- National Health and Medical Research Council (2018). *Australian Guide for Responsible Conduct of Research*.
- Netherlands Code of Conduct for Research Integrity (2018). <https://doi.org/10.17026/dans-2cj-nvwu>.

- Reason, P. & Bradbury-Huang H. (eds.) (2008). *The SAGE Handbook of Action Research: Participative Inquiry and Practice*. Second Edition. SAGE Publications.
- UK Research Integrity Office. (2023). *Code of Practice for Research*.  
<https://doi.org/10.37672/UKRIO.2023.04.codeofpractice>.
- UNESCO. (2022). *Recommendation on the Ethics of Artificial Intelligence*, adopted on 23 November 2021. United Nations Educational, Scientific and Cultural Organization.
- Van Buggenhout M., Christiaens J. and Gutwirth S. (2019). *Promoting Integrity as an Integral Dimension of Excellence in Research. Final Report on the Incidence of Misconduct* (PRINTEGER).
- Wimmer, R. D. & Dominick, J. R. (2014). *Mass Media Research: An Introduction*, Tenth Edition, Cengage Learning.



## Health and Identity in between the Private and the Institutional. The Case of Catherine, the Princess of Wales

Ramona-Gabriela HOSU<sup>1</sup>, Sînziana JURĂU<sup>2</sup>

**Abstract.** The singular objective of the present study is the examination of the relationship between private identity and institutional identity, labelled as 'body natural' & 'body politic' (Ernst Kantorowicz), with regard to Catherine Middleton's oncologic experience in the period January 2024 - January 2025, as publicly announced by herself or by the Palace. With regard to the tense, dynamic and complex structural duality 'private/institutional', the subject becomes more visible than usual because illness makes this duality more conflicting: an ailing body is by definition natural and private, but it is the political/institutional body that has to 'communicate' it, to make it public. The focus is on seven press releases that are to be interpreted as an identity management act, attempting to see which of the two identities is speaking, which is visible, which is dominant, which is protected by being back staged, which is used to benefit the other.

**Keywords:** private vs. institutional identity, body natural vs. body politic, health, identity management, advocacy, communication strategies, reception

---

<sup>1</sup> Associate Professor, PhD, Habil., Department of Journalism and Digital Media, Babes-Bolyai University, Cluj-Napoca, Romania, ramona.hosu@ubbcluj.ro

<sup>2</sup> Lecturer, PhD., Department of Journalism and Digital Media, Babes-Bolyai University, Cluj-Napoca, Romania, sinziana.jurau@ubbcluj.ro



## 1. Introduction

“What, then, is the need for a *further* debate about ‘identity’? Who needs it?”, asks Stuart Hall in a study published in *Questions of Cultural Identity* (2003, p. 1-17). The fact that identity is not a given, a stable ‘entity’ one is born with, but a fluid and dynamic process constructed in and through discourse has turned into a cliché in more or less recent years, in our post-postmodern or digimodern era dominated by accelerated globalization and exacerbated ICT development. Today’s social media and global networking mean hyper-connection and hyper-communication, in which the physical and digital realities are merged and the line between the private and the public gets indistinct, which problematizes the sense of self and belonging, and which consequently results in a conflicting type of identity or an identity crisis. Hence, to quote Stuart Hall again, the ‘necessity of identities’ should be “unambiguously acknowledged” because “the question, and the theoretization, of identity is a matter of considerable political significance” (p. 16).

The dichotomy private identity/ institutional identity makes the subject of the present study. The interplay, tension and conflict between the two ‘realms’ is encompassed in the term ‘body politic’ – as the “ancient metaphor by which a state, society, or church and its institutions are conceived of as a biological (usually human) body”, according to *Encyclopedia Britannica* (Rollo-Koster, 2017). The term has ancient origins, and Plato refers to it in his *Republic* when defining the state as a human body endowed with reason, spirit and desire that must function conjointly in harmony. In 1957, in *The King’s Two Bodies. A Study in Medieval Political Theology*, Ernst H. Kantorowicz reinterprets the metaphor of the state as a living organism making the distinction between ‘body natural’ and ‘body politic’ while analysing this structural duality in the figure of the King. Recently, on 27 March, 2024, *The New York Review* publishes an article titled “Kate’s Two Bodies” in which the author, Frances Wilson, journalistically elaborates on the duality mortal-immortal (i.e. natural vs. political body) when discussing a PR crisis and Kate Middleton’s cancer diagnosis. The article concludes: “Until she can reverse her Faustian pact and reclaim her Body natural, she will remain on the bench reciting prepared statements, alone of all her sex”, therefore finding that there are no traces of the ‘natural body’ in the official statements coming from the Palace because Kate “traded her Body natural” “for a Body politic”, seeing it “at the heart of the mess the palace PR machine has got itself into”. (Wilson, 2024)

It is known that the Princess of Wales's health issues at the beginning of 2024 and the press releases related to them generated huge public controversies - starting from the lack of transparency related to her abdominal surgery in January 2024, followed by an altered family photo and then by a video announcement of the diagnosis. The so-far mentioned structural duality 'natural-political' body is evidently the case here: it is the disease which makes the conflicting duality visible. Between January 2024 and January 2025, there were seven press releases of the British royal family in relation to the Princess of Wales's oncologic journey and the relationship between private identity and institutional identity as represented in these seven texts is the singular objective of the present paper. Each press release is to be interpreted as an identity management act, attempting to see which of the two identities is speaking, which is visible, which is dominant, which is protected by being back staged, which is used to benefit the other.

## **2. Private versus Institutional or 'Body Natural' versus 'Body Politic': some theoretical considerations**

There are five theories that are brought together here to outline the conceptual framework: Stuart Hall's *Who Needs Identity?* (1997); Ernst H. Kantorowicz' *The King's Two Bodies* (1957); Erving Goffman's *The Presentation of the Self in Everyday Life* (1959); Sandra Petronio's *Boundaries of Privacy. Dialectics of Disclosure* (2002), and Stuart Hall's *Encoding and Decoding in the Television Discourse* (1973). They are heterogeneous theoretical contributions: one is a cultural theory, the second one is a medieval history of institutions, the third one is interactionist sociology, the fourth is an interpersonal communication theory and the fifth one views the dialogue of the media content creator and the public. Despite being incongruous, they all centre on the analysis of identity construction and communication by addressing the same question: how is identity constructed in relation to the other(s), institutions included here. Hall explains why identity is not fixed but it entails discursive strategic positioning. Kantorowicz points to the conversion of body natural/ private identity into body politic as institutional entity. Goffman refers to the mechanisms of presentation of the Self. Petronio focuses on the management of the boundaries between what it is shown and what it is hidden in an identity. Hall's encoding/decoding theory outlines the complex processes of media message interpretation and transformation in the circuit connecting content production and content reception.

To begin with the first one: identity in itself, explains Hall, occurs in the “relationship between subjects and discursive practices” (2003, p. 2) and this leads further on to the concept of identification, which corroborates the processuality of identity construction as a changeable, unstable hybrid that intersects linguistic, social, cultural and historical influences. Thus, identities are articulated by discursive practices and discursive positionality, that is, to quote Hall again: identification “is lodged in contingency”, it “does not obliterate difference”, it is a signifying practice that is “subject to the ‘play’, of *différence*” and it “obeys the logic of more-than-one” (2003, p. 3). In other words, identity is relational and it is constructed by what it excludes. Criticizing essentialism, Hall sustains that there is no primordial, authentic, inherent Self hidden in social strata but this is rather produced in and through *representation*, in a meeting point of “*suture*” between the discourses that “interpellate” us and the “processes which produce subjectivities” (2003, p. 5-6). Identity is therefore a relational discourse neverendingly constructed by and in contexts of power.

*The King's Two Bodies* of Ernst H. Kantorowicz (1957) as a study of medieval political theology uses the classical formula ‘The King is dead. Long live the King’ in order to make the distinction between the king’s natural body as mortal, vulnerable, unhealthy and his political body as immortal and incorporeal. Kantorowicz uses English jurisdiction in order to explain the difference: «“[...] to the natural Body [of the king] there is conjoined his Body politic which contains his royal Estate and Dignity,” or that “the Body politic, annexed to his Body natural, takes away the Imbecility of the Body natural.”» (Kantorowicz, 1957, p. 494). The corporeal structural duality of the king’s identity makes a unit, eventually, but the two bodies are analytically separable. Dignity in ‘*dignitas non moritur*’ is the juridical formula used by the German historian in order to refer to the modern idea of ‘institution’ as a separate entity in one’s identity. To sum up the perspective: ‘The Crown as Fiction’ (p. 336) is representative for the political body but even if it is ‘fiction’, it does not ‘lie’ because it solves a real problem by living beyond the biological and surpassing the fragility of those it represents. The ‘doubleness’ of one’s body (natural / political) in one entity is a tense condition that generates representation crisis.

Erving Goffman’s *The Presentation of the Self in Everyday Life* (1959) is the well-known canonical text that introduces the idea of the dramaturgical pattern by seeing human social interaction as theatrical performance in

which individuals manage impressions for the public. The self is not some well located inherent essence to be externalised but rather it is “a dramatic effect arising diffusely from a scene that is presented, and the characteristic issue, the crucial concern, is whether it will be credited or discredited” (Goffman, 1959, p. 252-253). In his view, there is the “front region” and the “back region” and a team on stage “from which the performed character’s self will emerge” coupled with another team, the audience “whose interpretive activity will be necessary for this emergence” (p. 253). Important in this arrangement is the duality of the two ‘regions’ because this is the condition for impression occurrence and impression management. Authenticity in the matter of self construction is not seen as a moral category but rather a matter of information dialogue and management in between front stage and back stage. One fundamental aspect in line with the present study is Goffman’s idea of the two types of communication that he calls “expressions given” and “expressions given off”, where the latter is unintentional even if communication is “purposely engineered or not” (Goffman, p. 4) – the important implications of what becomes unintentionally transparent in one’s ‘expression’, even if back staged or hidden.

The fourth theoretical approach to follow here is Sandra Petronio’s *Boundaries of Privacy. Dialectics of Disclosure* (2002). Her theory of Communication Privacy Management (CPM) sees the management of private information as a systematic process in which the tension between what is disclosed and what is concealed is regulated by people through some boundaries defined by three management processes: (1) privacy ownership, (2) boundary coordination – rules negotiated culturally, contextually and relationally and that include personal and collective boundaries, hence the importance of co-ownership of private information and the need to negotiate the shared set of rules, and (3) boundary turbulence as the case when the shared rules are infringed, resulting in conflict, distrust, the need for renegotiation. Disclosure and concealment coexist in interplay, Petronio states, and they are tense but not opposed. Moreover, there is no ‘natural’ and complete transparency or secrecy when it comes to private information. In Sandra Petronio’s words: “In order to fully grasp the nature of private disclosures, we not only have to consider the individual who is revealing or concealing, but we also must focus on how the decision affects other people [...] CPM assumes that others are also central to discerning the tension between being public and private. CPM uses the metaphor of boundaries to illustrate that, although there may be

a flow of private information to others, borders mark ownership lines so issues of control are clearly understood. Thus, regulating boundary openness and closedness contributes to balancing the publicness or privacy of individuals.” (Petronio, 2002, p. 3).

As already mentioned, the four theories above intersect in focusing on the dialogue and distinction between the public/ political/ institutional/ front staged individual *and* the private/ natural/back staged individual, more specifically: in the case of Catherine Middleton’s cancer journey, health issues and medical information become the negotiated co-property of the patient, the institution and the public. Each of the seven official press releases (in 2024 and the beginning of 2025), which make the object of study and analysis in what follows here, is to define and redefine the boundaries of both her private identity and institutional identity, as acts of negotiation of what of the private becomes public/ institutional and vice versa.

Stuart Hall’s reception theory, stated in his 1973 paper titled *Encoding and Decoding in the Television Discourse*, argues that the communication process should be seen as a fragmented circuit, where a miss-alignment is frequently encountered between the communication intent of the media producer and the interpretation given by the audience to the message. As consumers decode the messages, they use different interpretive frames, derived from their own cultural, economic and social backgrounds. Hall offered an over-view of the predictable decoding perspectives the audiences could adopt, identifying the dominant-hegemonic position, the negotiated position and the opposition one. Adopting the dominant position is ideal in communication situations, whereas the audience accepts and decodes the message and its codes as they were initially intended by the producer, thereby resulting in a perfect communication process. However, more realistic communication situations recognize the audience critically positioning itself either by negotiating its position by accepting the legitimacy of the message on a general level but also by assigning it a unique interpretation, or by adopting a clear opposition status, the case in which the audience actively reconstructs an alternative, a critical interpretation of the otherwise perfectly understood intended meaning and preferred cultural codes.

### **3. Media reporting and the right to private and family life in the UK. A turbulent recent history**

#### ***3.1 The European Convention on Human Rights as an external guarantee of the right to private family life***

The death of Diana, the Princess of Wales, in 1997 shocked the world and it focused the moral and legal scholarly lenses on the French, British and ultimately European legislation regarding the right to privacy of public figures and its delicate balance with freedom of expression of the media.

Although the UK signed the European Convention on Human Rights (ECHR) in Rome in 1950 and ratified it in 1951, due to the country's dualist approach to international law, the ECHR remained an external and illusive point of reference for national judges, when exercising their discretion in case of ambiguity between the national and international legislation. It was only in 1998 that The Human Rights Act expressly incorporated the convention into UK law, giving effect to the rights and freedoms guaranteed under the **European Convention on Human Rights (ECHR)**, including the rights comprised in articles 8 and 10 regarding the right to private life, home and correspondence and the right to freedom of expression within the UK legal system. Prior to this adoption, British citizens could not rely on domestic courts' capacity of enforcing the right to privacy, thus being forced to seek the recognition of this fundamental right by ratified international treaties and the international courts as venues providing the liability and enforcement mechanisms (Deringer, 2003, p. 208-211).

A testament to a tradition of fierce debate and probably fundamentally opposite views of what the right to privacy covers and includes, three of the nine photographers initially accused of manslaughter in the case were convicted to pay a symbolic fine of one euro each for intrusion into privacy by the French courts. The French debate on whether the interior of the car is still considered a private space after the accident forced it open span for almost a decade, being presented in front of multiple judges and courts prior to concluding that "by showing Dodi Al-Fayed mortally injured, bloody, his body broken by the shock of the accident, the photographers unquestionably committed a breach of the respect due to a human being and as a result violated the personal privacy of the individual concerned" (referred to in *Al Fayed v France*, Application no. 38501/02).

### ***3.2 The European Court of Human Rights and the limits of private life of royalty in the Von Hannover cases. Definitions and balancing criteria***

Similarly harassed by paparazzi in daily contexts, the eldest daughter of Rainier III, Prince of Monaco and Grace Kelly, Princess Caroline of Monaco, led a long-lasting legal fight for her right to privacy, as well as in support of her, and by extension, all “figures of contemporary society par excellence” as they will be defined by the judges of the Courts, to be able to freely enjoy their life in public (spaces and spheres) without the self-limitations imposed by the presumption that, even if not exercising public office or attending public functions, their private lives would still be subjected to public scrutiny through the ruthless lenses of the media and paparazzi.

The cases brought by Princess von Hannover, initially in front of the diligent and thorough German courts, and later at the European Court of Human Rights, offered the European court the possibility to contribute to the creation of a long-lasting European framework of accountability of the media for intrusion into privacy, along with guidelines and criteria for the national judges of all signatory States of the Convention to be used for determining and striking a just balance between the right to respect private life as provided by Article 8 of the Convention and freedom of expression and its limited exercise, as stipulated by Article 10 of the Convention, especially in the context of photographs of public figures taken in daily, non-official settings.

In *Von Hannover v. Germany* 1 (2004), the Court concluded that the Princess’ right to privacy was violated, as German courts failed to ascertain the legal framework necessary to offer the Princess sufficient protection. The European Court noted and emphasized that photographs published in the tabloid press are “often taken in a climate of continual harassment” and offered as a decisive criterion for the necessary balancing act that judges (national and international) need to perform between the free enjoyment of both rights - the “contribution to a debate of general interest”. The reasoning of the German courts regarding the “relative public figure” and “spatial isolation” criteria were rejected by the European Court as excessively vague, as it finally argued that the legitimate expectation of protection of private life is enshrined in the ECHR, irrespective of the status of the person requesting protection.

Von Hannover v. Germany 2 (2012), was decided by the Grand Chamber. The Court found no violation of Article 8 of the convention pertaining to the right to privacy, noting the fact that the German courts have successfully followed the recommendations made by the Court in its 2004 judgement and thus successfully reached a fair balance between competing rights, whilst exercising their margin of appreciation. The Grand Chamber took this second opportunity to introduce a clear structure in the balancing exercise performed by courts, distinguishing between five criteria, namely:” (i) the contribution made to a debate of general interest, an interest not confined to political or criminal matters; (ii) how well known the person concerned is and the subject of the report; (iii) the prior conduct of the person concerned, including any previous cooperation with the press; (iv) the content, form and consequences of the publication; and (v) the circumstances in which the photographs were taken, including consent and the manner of obtaining them”. Upon using the above-stated criteria, the Court concluded that the photograph accompanying an article on Prince Rainier's illness was related to an event of contemporary and general interest, and that the domestic courts had adequately weighed the criteria, thereby remaining within the margin of appreciation to be afforded to national authorities.

#### **4. Case Study: The Private & the Institutional Catherine Middleton**

##### ***4.1. Context***

Between January 2024 and January 2025, history registered unprecedented news regarding the British royal family, about two cancer diagnoses, treatments and healings, of King Charles and of the Princess of Wales, a turbulent relationship of the royals and the public, including the press, and reduced public engagements. In January 2024 we found out about King Charles’ prostate cancer, followed by intensive treatments and no duties performed publicly but only behind closed doors, plus the same month recorded the planned abdominal surgery of Catherine Middleton, with no specific diagnosis mentioned. This was followed by absence from public duties and intensive speculations in the media till March 2024, when an altered photography was posted on the official social media accounts, followed by kill notices from major news agencies, and intensive speculation

about the princess' real health condition, controversies about transparency and trust in the royal family, as well as debates about misinformation, digital manipulation and reality – known as 'Kategate'. On the official account on X, the princess later apologised by mentioning her amateur skills in photo editing. After the manipulated photo, a private video was released announcing cancer and some details about what happened post-surgery with regard to health and family matters. Later, the public appearances of Princess Kate at Trooping the Colour, in June 2024, and at Wimbledon, in July 2024, were hugely acclaimed and the private family video released in September 2024 eventually announced the end of chemotherapy. In January 2025, Princess Kate announced cancer remission. Prince William continued his public activities but called 2024 a 'brutal' one, while trying to perform his duties and support his family. Good news marked the end of 2024, with Princess Beatrice's pregnancy news, and in January 2025, the public found out there were plans for the King's and Princess Kate's gradual return to regular public duties.

#### ***4.2. Research Objective and Research Questions***

As previously mentioned, there is one objective in the present study: the examination of the relationship between private identity and institutional identity, labelled above as 'body natural' & 'body politic' (Ernst Kantorowicz), with regard to Catherine Middleton's oncologic experience in the period January 2024- January 2025, as publicly announced by herself or by the Palace. With regard to the tense, dynamic and complex structural duality 'private/institutional', the subject becomes more visible than usual because illness makes this duality more conflicting: an ailing body is by definition natural and private, but it is the political/institutional body that has to 'communicate' it, to make it public.

Each media message is to be seen as an act of identity construction *and* identity management in strict connection with the dichotomy private/institutional, focusing on delineating which of the two identities is expressed, which is visible or hidden and how it is revealed or concealed, which is dominant, which is used to benefit the other, what changes occur in the communication of the two identities over a year, since the beginning of the health issue till the complete clinical and personal recovery. Hence, the research questions (RQ) are the following:

RQ1 What type of identity is dominant in each text and what makes it dominant?

RQ2 How does the relationship private – institutional identity (or ‘body natural’ – ‘body politic’) change chronologically, from the first to the last official press release?

Building on Stuart Hall’s reception theory, with the aim of studying how the prominent top six British online news outlets have received and represented the public official statements made by the Kensington Palace or by the Princess of Wales, we complement the initial analysis with the following research questions:

RQ3 How do media outlets interpret and present the official statements, considering the initially intended dominant identity?

RQ4: To what extent do the media outlets respect the right to privacy of the Princess of Wales and her family when reporting on her health struggles?

### ***4.3. Sampling. Data Collection & Analysis***

In strict connection with the first two research questions, because the focus in this case is on analysing the underlying meanings and themes related to the topic, as they appear in media messages, QCA/ Qualitative Content Analysis is the method to use. Words, images and videos are to be categorized into specific thematic patterns/ codes that define both identity construction and identity management – for the two types of identity, moving beyond numbers and frequency in the specific media content, therefore beyond QnCA. Since it starts from the concepts and theories presented above – that contribute to identifying the coding categories, this is to be a deductive approach that implies systematic reading and/ or watching of the material to identify the codes and possible broader themes - all done manually.

For RQ1 and RQ2, the Corpus of analysis consists in seven official public announcements/ press releases made by Kensington Palace or by the Princess of Wales on official channels (videos released to the press or on official social media accounts) – all public and accessible through press archives (BBC, ITV, CNN, NBC), and available on *The official Instagram account of the Prince and Princess of Wales, based at Kensington Palace, @princeandprincessofwales*.

The materials have been manually and purposively selected for the period January 2024-January 2025, with a focus on identity matters in strict connection with Catherine's health issues, more specifically Catherine Middleton's oncologic journey, from the first official announcement of the health problem, namely the abdominal surgery, to the official announcement of cancer remission. Consequently, there have been identified seven media materials codified as U1, U2, U3, U4, U5, U6, U7, namely:

U1 = The Kensington Palace official statement, 17 January, 2024: the planned abdominal surgery;

U2 = The Kensington Palace official statement, 29 January, 2024: the return home and the progress in recovery;

U3 = Mother's Day manipulated photo + caption, 10 March, 2024;

U4 = The diagnosis disclosure video, 22 March, 2024;

U5 = A photo + caption announcing progress and the challenges of chemotherapy, 14 June, 2024;

U6 = The royal family video + caption announcing the end of preventive chemotherapy, 9 September, 2024;

U7 = A photo + caption about a public visit at The Royal Marsden, 14 January, 2025, announcing remission.

The unit of analysis for each of the seven media materials is the complete 'text' released to the public (the official written announcement, the posts – photos, videos and captions). There are five dimensions / codes to be used for gathering and analysing data, generated by the conceptual instruments presented in the theoretical part of the present study, as follows:

C1 = Medical information and level of specificity - Kantorowicz's concepts. This dimension is to see how private matters (aspects related to health, i.e. 'body natural') are made public (by 'body political'). 0. no medical information; 1. generic information; 2. treatment but no diagnosis; 3. specific diagnosis.

C2 = (Non)Disclosure Explanation - Goffman's and Petronio's theories; this is meant to justify the 'performance' and the exchange between front stage / back stage, revealed / concealed. 0. absent; 1. prevention of speculation; 2. protection of family / children; 3. advocacy; 4. privacy ownership.

C3 = Speaker & Dominant Identity Type - the concepts of Kantorowicz, Goffman and Petronio. This dimension points to the 'voice' speaking and signing the message, as follows: S1. the institution (the princess is addressed in the 3<sup>rd</sup> person singular); S2. the subject (in the 1<sup>st</sup> person singular) and S3. both / mixed; at the same time, it includes the type of 'body' that is expressed

as dominant; T1. body natural (as patient but also as mother and wife, considering familial status as part of Kantorowicz' 'natural' body; T2. body political. These two sub-categories (speaker & type of identity) are seen together because they both operationalize the duality private/ institutional.

C4 = Advocacy - in connection with the concepts of Kantorowicz, Goffman and Petronio; this code is meant to establish the level of connection between the person and the public; advocacy is the point of conversion of private identity into a capital of institutional identity; 0. absent; 1. expressed discursive solidarity; 2. institutional advocacy (patronage).

C5 = Pressure in connection with Sandra Petronio's borderline turbulence; the degree of pressure registered when the 'text' was released, in relation to the transparency of information and public speculation; 1. low; 2. moderate; 3. high.

In order to answer RQ3 and discover how media outlets interpret and present the official statements, considering the initially intended dominant identity, we suggest transforming the potential reception values based on Hall's theory into the following indicators:

C6 = Type of reception; 0 = conforming reception by the media, whereby the studied media outlet accepts the original intended meaning of the official statement and reproduces it with minimal challenge or reinterpretation; indicators of this framing include quoting official sources without additional critique; 1= negotiated reception, whereby the studied media outlet presents parts of the official narrative but contextualizes or questions aspects of it by including expert commentary and underlining missing information or opaque answers from the official sources; 2 = oppositional reception, whereby the news outlets reject the original meaning by creating an alternative interpretation. In order to distinguish between the negotiated reception previously introduced and coded with 1, the oppositional reception by the media will be attributed to the authors and articles that focus on controversies and contradictions, interpreting the official statements as examples of strategic communication.

In order to answer RQ4 related to the respect afforded by the analysed media outlets regarding the right to privacy of the Princess of Wales and her family when reporting on her health struggles, the indicators were derived from the criteria suggested by the ECtHR in its first and second Von Hannover v. Germany cases, surveying whether the article includes a public interest point of view, by referring to either office or official functions, and whether

the article respects the privacy requests made by the Princess of Wales, in the interest of protecting her family life and health related data privacy. Throughout the analysis, this indicator was marked with C7 for consistency.

The adequate research methods for answering RQ3 and RQ4 consist in performing qualitative and quantitative content analysis of the online articles published by the online media corresponding to the previously identified key moments. For this purpose, a corpus of 75 articles published online on the following websites have been manually selected and analyzed: [bbc.co.uk](http://bbc.co.uk), [theguardian.com](http://theguardian.com), [reuters.com](http://reuters.com), [mirror.co.uk](http://mirror.co.uk), [thesun.co.uk](http://thesun.co.uk) and [dailymail.com](http://dailymail.com). The articles were selected by using the available search engines on the media outlets' websites, introducing the specific dates of the events or press-releases as well as the name of the Princess of Wales or Catherine. For the numerous cases in which media reporting spanned over several dates, the articles were included in the sample for analysis if they presented either significant new developments or if they had the potential to align with any of the three potential types of reception, as derived from Hall's theory, or if they had the potential to illustrate a criteria recommended by the Von Hannover 2 jurisprudence.

#### 4.4. Results. Discussion

##### 4.4.1. Private and Institutional Identity in the official press releases of Catherine and/ or Kensington Palace – RQ1 and RQ2

With regard to RQ1 and RQ2, the Analysis Grid below synthetically presents the key findings for each dimension / code described above:

**Table 1.** Analysis grid

| Corpus Units | C1<br>Medical<br>Information<br>Specificity | C2<br>(Non)Disclosure<br>Explanation | C3<br>Speaker &<br>Identity Type<br>(dominant type<br>in bold) | C4<br>Advocacy | C5<br>Pressure |
|--------------|---------------------------------------------|--------------------------------------|----------------------------------------------------------------|----------------|----------------|
| U1           | 1 + 2                                       | 1 + 2 + 4                            | S1 + T1 + <b>T2</b>                                            | 0              | 1              |
| U2           | 1 + 2                                       | 3                                    | S1 + T1 + <b>T2</b>                                            | 1              | 1              |
| U3           | 0                                           | 0                                    | S2 + T1 + T2                                                   | 0              | 3              |
| U4           | 2                                           | 1 + 2 + 3 + 4                        | S2 + <b>T1</b> + T2                                            | 1              | 3              |
| U5           | 2                                           | 0                                    | S2 + <b>T1</b> + T2                                            | 1              | 1              |
| U6           | 2                                           | 2                                    | S2 + <b>T1</b> + T2                                            | 1              | 1              |
| U7           | 2                                           | 3                                    | S3 + <b>T1</b> + <b>T2</b>                                     | 2              | 1              |

The data in the table above allow for a discussion on identity construction and management that follows a transformative cyclic trajectory that chronologically corresponds to the seven ‘texts’ in the corpus of analysis.

The first two press releases (the official statements of Kensington Palace announcing Catherine’s abdominal surgery, with no specific diagnosis, and her return home) stand for the first phase, called here ‘*controlled institutional identity*’. It is the institution that talks front stage, that is dominant; the princess is mentioned with her full title and in the 3<sup>rd</sup> person singular; the return to public duties is announced for after Easter; the justification for non disclosure of medical details is linked with prevention of speculation, with children protection and there is explicit reference to her privacy rights; advocacy stands in the mentioning of The London Clinic, in the second text, which also expresses discursive solidarity. The private/natural body is invisible, back staged, to use Ervin Goffman’s terms. The ailing body that underwent surgery, the recovery and the family are protected by explicit borderline rules and the privacy rights formulated by the institution (Sandra Petronio’s concepts).

The problem with this type of ‘controlled institutional identity’ is not to be related to media content construction and identity representation but rather with communication strategy. The institution presupposed that the public would respect the borderline between ‘front’ and ‘back’ [Goffman and Petronio’s terms], seeing ‘back stage’ (the private) as a forbidden region. Instead, the social media of the year 2024 formulated the question ‘Where is Kate?’ as if collectively asking for access to the private individual or to back stage.

Unit 3 or the manipulated Mother’s Day photo corresponds to the second stage of identity representation. The table above does not identify any dominant type of identity here, of the two types, and this is because despite the fact that it is Kate that speaks in the caption and that signs the message with ‘C’, directly addressing the public and thanking for the support, the altering of the picture casts doubt on the credibility of the message itself, on who the real person addressing the public is. Considering the impact and the aftermath of the manipulated picture, this is the moment when the relationship between private/ institutional gets out of control. The private is moved front stage next to the institutional and the honesty and credibility of the two bodies, natural and political, are questioned. Kate’s

apologies issued later on X are to be seen as a private and strategic positioning front stage meant to correct an institutional error. Evidently, there is boundary turbulence here, Petronio's concept, and high pressure resulting from intensive speculation in the dialogue of the person (private and institutional Kate) with the public. This second phase in the management of identity could be called 'diffused identity' – because the line between the private and the institutional body gets blurred.

The next three units of analysis, U4, U5 and U6 correspond to the third stage of identity representation: the '*assumed private identity*', in which case the private is front staged and the institutional gets back staged – an exchange of positions and of roles. The famous diagnosis disclosure video belongs to the private individual that restates the reasons for the initial concealment of the health condition (those mentioned in the first Kensington Palace statement), namely: prevention of speculation, protection of family, privacy ownership. It is 'body natural' which dominates the dichotomy in these three texts: the 1<sup>st</sup> person singular speaking, in a natural setting, using words in the very emotional register. The performance (Goffman's term) is controlled by the source and the public eventually gets access to the 'private'. The initial confidentiality borderline (the ten weeks of silence) is explained by a very humane reason: the need to protect the children and the family, which can rarely be contested (Petronio's 'boundary coordination rules'). The disclosure in U4, 5 and 6 emphasizes privacy as intimate discourse not as intimate information. It is the case of the private individual that represents and constructs the 'body', not the institution.

Phase four of identity construction and management is registered in the last text, U7. The private experience of disease is turned into a resource for institutional identity. The news about the hospital that treated her, the Royal Marsden, the voice that speaks (1<sup>st</sup> person singular, the patient, but her new institutional role as Joint Patron of the hospital, as well) are suggestive for 'converted identity': a type of identity that successfully co-owns the natural and the political, as a result of transforming private experience into an institutional good. Illness is the instrument for advocacy. Thus, private identity or 'body natural' legitimizes the new institutional identity of Catherine, the Princess of Wales. In other words, the relationship private / institutional is no longer dichotomised because it gets new meanings that reinstate the duality in harmony.

A definite transversal finding is that the specific diagnosis (the type of cancer) is never disclosed. This precise frontier between the private and the institutional is never crossed.

#### 4.4.2. Media reception of the official press releases of Catherine, the Princess of Wales, and/ or of Kensington Palace – RQ3 and RQ4

##### 4.4.2.1. Media framing of Catherine's hospitalisation

The media's reception and decoding of the initial messages regarding the Princess of Wales's hospitalisation, in an initial corpus consisting of the six articles published by the six selected media outlets on the same day, were analyzed through the prism of the codes C1-C6. As previously observed, the initial official statements issued by Kensington Palace, marking *the controlled institutional identity* phase offer a strong initial emphasis on both the right and the desire for privacy strongly correlated, also from the emotional standpoint, with the desire to enhance children's protection.

The table below summarizes, per media outlet, the outcome of the qualitative analysis performed.

**Table 2.** Qualitative analysis

| Outlet       | C1  | C2    | C3          | C4 | C5 | C6 |
|--------------|-----|-------|-------------|----|----|----|
| BBC          | 2   | 1,4   | S1 / T2     | 0  | 1  | 0  |
| Reuters      | 2   | 1,4   | S1 / T2     | 0  | 1  | 0  |
| The Guardian | 2   | 1,2,4 | S1 / T2(T1) | 0  | 1  | 0  |
| Mirror       | 2   | 1,4   | S1 / T2     | 0  | 2  | 1  |
| The Sun      | 2   | 1,4   | S1 / T2     | 1  | 2  | 1  |
| Daily Mail   | 1,2 | 1     | S3 / T1,T2  | 1  | 2  | 1  |

There are several constants in the values the codes register. In terms of presenting the information with medical specificity, as well as in terms of conveying to the audiences the non-disclosure rationale, the studied media outlets present a consistent pattern. By borrowing the exact institutional choice of words and tone, the media uniformly discloses the *abdominal surgery*, framing it as *successful*. The rationale supporting the limited disclosure is equally uniformly shared, conveying the initial institutional requests of refraining from speculation while exercising fully the right to privacy with the scope of protecting the family and children. We observe

thus Petronio's limits assumed proactively. Similarly, we observe the same institutional dominant speaker when reading the articles published by five out of six of the media outlets, with *Daily Mail* being singular in its brief insertion of a third-party patient testimonial, stimulating, albeit briefly, a perception of Catherine as a vulnerable natural body.

Advocacy perspectives were not registered explicitly, accurately reflecting the initial institutional code. The two coded values observed upon analysing *The Sun* and *Daily Mail* were recorded taking into account the sympathetic tone of the reporting.

The fifth code pertaining to a potential pressure exercised on the limits, as defined by Petronio, were low to moderate. The association of words *shock* and *health*, as well as the narrativization of the hospital environment found in the *Daily Mail* reporting were recorded as moderate and without breaching the set boundaries.

We can observe and conclude that, with regard to Hall's decoding categories, *BBC*, *Reuters* and *The Guardian* quoted the institutional statement with minimal to no interpretation, thus conforming to its initial intended reading of the Princess of Wales as an institutional body performing a dignified and justified temporary withdrawal from future public functions. *The Sun*, *Mirror* and *Daily Mail* were registered as presenting their audience with a negotiated reception, presenting third-party insider narratives of the hospital environment of by reframing the news enhancing emotional perspective (*The Sun*).

In terms of respecting the legal boundaries by fulfilling the criteria stipulated by the ECtHR in *Von Hannover v. Germany*, the whole corpus analysed complies with the scope of the provisions of articles 8 and 10 of the European Convention on Human Rights. The media institutions that positioned themselves in the conforming category, namely *BBC*, *Reuters* and *The Guardian* fulfil almost automatically the criteria stipulated by the ECtHR in *Von Hannover 2*, since the included perspective is that of voluntary limited disclosure, with the scope of justifying the temporary withdrawal from the performance of the institutional functions. The media institutions that positioned themselves in the negotiated reception category are equally compliant but illustrate more closely the debate sparked and settled in *Von Hannover 1* case regarding the potential interest of the audiences and the reasonable expectation of privacy a public person might hold. If and as long as no clear intrusion into privacy is performed, the media outlets remain compliant, as we see in this case.

A potentially interesting observation is that when the private information is voluntarily disclosed, and the limits of the particular disclosure are respected by the media outlets, whilst still presenting the audience with the information that responds to the public's justified public interest, media institutions will find themselves ECHR compliant – a simple and foreseeable potential “solution” to the intrusion into privacy conundrum, when clear public interest can be identified.

#### 4.4.2.2. Media framing of the controversy about Mother's Day photo

As previously discussed, this second stage of identity representation marks the turning point in which the relationship between the institutional and the private are no longer carefully balanced. The act of manipulation itself, as well as its consequences, cast doubt regarding the credibility of both the natural and the political body, and, in order to study the reception of this “diffused identity”, with privacy boundaries moved - if not removed - completely, we have identified, using the same selection method, twenty-three online articles, published by the same initially selected six media outlets, on the same dates. The monitored events that determined the inclusion into the corpus of analysis were the issue of Kensington Palace of the edited photograph, the speculations surrounding it, the withdrawal otherwise dubbed as a “kill order” by several news agencies, including *Reuters*, as well as the reports of Catherine, the Princess of Wales's apology and justification for the issuance of the edited photograph. The table below presents aggregated results per outlet and per code, registering the general tendency of the media outlet, for the purpose of presenting the results of the qualitative analysis performed.

**Table 3.** Aggregated results per outlet and per code

| Outlet              | C1 | C2    | C3    | C4 | C5 | C6  |
|---------------------|----|-------|-------|----|----|-----|
| <i>BBC</i>          | 2  | 1,2,4 | S3/T2 | 2  | 2  | 1   |
| <i>Reuters</i>      | 2  | 1,4   | S1/T2 | 0  | 1  | 0   |
| <i>The Guardian</i> | 2  | 1,4   | S1/T2 | 2  | 3  | 1   |
| <i>The Sun</i>      | 2  | 1,3   | S3/T1 | 1  | 3  | 1–2 |
| <i>Mirror</i>       | 2  | 1,2   | S3/T1 | 1  | 3  | 1   |
| <i>Daily Mail</i>   | 2  | 1     | S3/T1 | 1  | 3  | 2   |

In terms of disclosure of the medical data, the codes are consistently registered at an identical value. All media outlets continue to respect and reflect, in their reporting, the generic medical reason behind the treatment, without including the diagnosis that caused the initial press-release and the withdrawal of the Princess of Wales from the public eye. We observe in this regard the full compliance of the media outlets with the institutional original message, as well as with the previously set privacy boundaries.

The explanations for the non-disclosure of the information continue to be mainly the prevention of speculation, seconded by the original specific request for privacy and the protection of children. While *Reuters* and *The Guardian* underline the choice of non-disclosure as institutional and an exercise of privacy-ownership, *The Sun*, *Mirror* and *Daily Mail* include the non-disclosure choice as a narrative catalyst, underlining “mounting pressure to reveal the truth”.

The third code, created to identify and investigate the speaker and dominant identity type and aiming at classifying and attributing a main frame either as a natural body or body political, also yields a series of interesting results when used to perform the content analysis of the published materials. As previously observed during the Mother’s Day photography controversy, the natural and body political aspects collapse and reciprocally undermine each other’s credibility. The dominant body is unstable and incoherent across the studied outlets, shifting from the clear institutional body observed at *Reuters* and *The Guardian* to the natural body in *The Sun*, *Mirror* and *Daily Mail*. Forensic attention to various details of the image, regarding blurred hands for example, convert the visual representation of the body into evidence, while channelling the attention of the audience towards personal aspects (like, for example, personal meaningful jewellery). The pronoun used to name the Princess (3<sup>rd</sup> singular) by *Daily Mail* for example, places direct culpability, clearly no longer shared with institutional communication. Directly linked, the fourth code studying advocacy is seen discussed directly by the *BBC* and *The Guardian*, that used the controversy as an opportunity to debate institutional accountability.

The fifth code registering pressure in connection with Sandra Petronio’s border turbulence is one of the most interesting analysis dimensions in this particular case, especially in connection with the codes related to medical confidentiality and (non)disclosure rationale. Due to what becomes now as clear tensions created by the privacy requests, the pressure points move towards either institutional pressure or personal pressure depending

on the editorial choice. Illustrating moderate to high institutional pressure, the *BBC* published not only exculpatory findings but also reflexive perspectives over the “unfairness”, whereas *The Guardian* for example performs a specialist 20 anomaly verification and analysis. *Reuters* is, in this group, the original yet most conformist outlier, as it created the original structural pressure through the rare kill notice of the photograph, whilst rhetorically performing almost no pressure whatsoever. In a stark contrast, pressure and turbulence are at their highest registered levels in the texts published by *The Sun*, *Mirror* and *Daily Mail*. Though all media outlets convey the admission-driven escalation of Kategate, in these cases we see either a high personal pressure exercised via crowd forensics and explicitly questioning the credibility of future health updates as addressed by *The Sun*, or, on a more factual basis approach, a high personal pressure by means of enumerating the kill-notice counting as escalation meter, complemented by the confession framing.

The sixth code, used to study the media’s reception overall, also registered significant variations between the studied media outlets. An “extreme” yet clear-cut position is adopted by *Reuters*, which initially publishes the photo, illustrating clear conformity, but it manifests firm alignment with its own standards and principles, issuing an almost surgical kill-notice in the light of the discovery that the photograph has been altered. The institutional and personal message is not just received and conveyed but also denied immediately from spreading once the miss-alignment with their own institutional role is seen, rather than reframed or interpreted. *BBC*, *The Guardian* and *Mirror* positioned themselves in a negotiated reception. Each presented the initial message, but it re-contextualized it, either by expert analysis, as seen in the case of *The Guardian*, or by initiating a reflexive questioning of the media’s own fairness, as observed in the case of the *BBC*. A clear oppositional reception is seen performed by the *Daily Mail*, which uses the admission of editing as evidence of a cover-up, and this constructs an alternative narrative of culpability. While each of the media outlets points out the potential inconsistencies and alterations of the image, *The Sun* adopts a mixed negotiated and oppositional reception of the message, inclining however towards attributing values of strategic communication to the Princess’s confession and apology.

When studying the potential intrusion into privacy, we observe, once again, the potential “blind spot” of the Von Hannover criteria: because the Princess acknowledged being the author of the edited version of the photograph, her prior conduct is the main reason that enables all media outlets, regardless

of how they have positioned themselves, to respect the threshold of maintaining a disciplined and strictly observed balance between information of public interest and the right to privacy of the portrayed person(s). Even considering all these aspects, recalling the fact that Von Hannover balancing criteria also recommend ethically obtaining the information deemed of public interest, we noted *The Sun* in breach of the right to privacy, by publishing paparazzi images other UK broadcasters declined.

#### 4.4.2.3. Media coverage of the Princess of Wales's cancer announcement

The Princess of Wales's announcement was performed via a diagnosis disclosure video. The "body natural", as previously observed, is set in an environment controlled by the source. The Princess of Wales reinforces her initial reason of protecting her privacy as a means of protecting the children. In order to study the reception of this new shift, of "*assumed private identity*" by the selected media outlets, we have included in this new corpus of analysis a number of twenty-one online media products, published between 22<sup>nd</sup> and 23<sup>rd</sup> March 2024, corresponding to the date of the initial video and including the further debates and reframing by the media. The table below lists, as previously, the aggregated results of the analysis, per each media institution, to facilitate an over-all reading.

**Table 4.** Aggregated results of the analysis per media institution

| Outlet       | C1 | C2       | C3        | C4  | C5 | C6  | C7        |
|--------------|----|----------|-----------|-----|----|-----|-----------|
| BBC          | 2  | 1+2      | S2; T1    | 0/1 | 3  | 1   | Compliant |
| Reuters      | 2  | 1+2      | S2; T1    | 1   | 3  | 1   | Compliant |
| The Guardian | 2  | 1+2      | S2; T1→T2 | 1   | 3  | 1   | Compliant |
| The Sun      | 2  | 2 (+1)   | S3; T1→T2 | 1→2 | 3  | 0   | Compliant |
| Mirror       | 2  | 4 (+2)   | S3; T1    | 1   | 3  | 0/1 | Compliant |
| Daily Mail   | 2  | 3 (+2,4) | S3; T1→T2 | 2   | 3  | 0/1 | Compliant |

The analysis performed (based on the criteria of medical information disclosure and level of specificity) revealed, for all media outlets, the following predictable outcome: all studied articles included in the sample presented the previous surgery and the preventive chemotherapy, however the type, stage and prognosis of the cancer are withheld, and the Palace's anti-speculation instruction is transmitted verbatim by all six studied outlets. *BBC*, *Reuters* and *The Guardian* remain aligned in their disclosure

rationale, the published materials registering prevention of speculation and protection of family as values. *The Sun*, *Mirror* and *Daily Mail* diverge in their approach. *The Sun*'s own headlines contradict the assumed intent of children's protection as motive. *Daily Mail* frames duty and leadership, whereas *Mirror* is approaching the subject by underlining the exercise of privacy-ownership.

The third code, C3, aimed at discovering the speaker and dominant identity type, yield the most unpredictable results. *BBC*, *Reuters* and *The Guardian* registered and conveyed to their audiences the Princess of Wales's own, first-person voice, as present throughout the video, with the body natural present as patient and mother, and the institutional body clearly fading. *The Sun*, *Mirror* and *Daily Mail* were recorded as S3 values for including commentary and externally sourced experts into the narration. The study of the advocacy code revealed a clearly expressed discursive solidarity in *BBC*, *Reuters*, *The Guardian*, *The Sun* and *Mirror*. An initial attempt to convert the confession into institutional advocacy is registered in the case of *Daily Mail*. The pressure code noted with C5, aimed at studying the borderline turbulence, was scored as high. Although the tone of the media articles remains sympathetic, they all refer to the edited Mother's day photograph / "Kategate" as previous notable incidents.

As the video itself embodies the assumed private identity, during the study of its reception and decoding we have the opportunity to study the media outlets' framing of it, as well as, where appropriate, in conjecture with the past incidents and their attributed meaning. As the Princess of Wales uses her own voice, and a calm yet emotional frame, reminding the reasons behind previous choices, namely the protection of children and of the "young family", the viewer is introduced to a narrative of owned privacy that reinforces and reinstates the boundaries. *Reuters* and *The Guardian* convey the message from a negotiated position, questioning the timing of the institution, rather than the person or the message. *The Sun*, *Mirror* and *Daily Mail* revive the institutional body by introducing secondary meaning via expert commentary and various formulas. As a consequence of their negotiated positioning, we see the institutional body revitalized and the body natural in a supporting role, framed as performing a valiant act of duty rather than a personal communication on a delicate theme. *The Mirror*'s article, offering an intrusive non-verbal communication analysis and attributing it meaning and feelings, represents one of the clearest examples of reconstruction of meaning, rather than a negotiation.

In terms of the legal perspective of the potential intrusion into privacy, the initial articles presenting the message exhibit clear respect towards the re-established and reinforced boundaries of privacy. *BBC*, *Reuters* and *The Guardian* use the moment as an opportunity to reflect on previous media failures, whereas the *Daily Mail* is the only media outlet outwardly recognizing the tension that this “compelled disclosure” exhibits and creates.

#### 4.4.2.4. Media framing of the Princess of Wales’s cancer update (14 June 2024)

A corpus of eleven articles published by the six media outlets was selected in order to study the reception of the ‘body natural’ made public. The selected articles conveyed or debated the written statement released by Kensington Palace on the evening of 14 June 2024, in which the Princess of Wales confessed that she was making “good progress” through ongoing chemotherapy, had “good days and bad days”, and would attend Trooping the Colour event the following day, thus officially announcing her first public appearance since the diagnosis announcement.

Similarly to the previous qualitative analysis, the same analytical criteria were used and are presented below in an aggregated table, facilitating an over-all perspective of the media outlets’ approach to the topic.

**Table 5.** Aggregated results of the analysis per media institution (synthesis)

| Outlet       | C1 | C2 | C3      | C4 | C5 | C6 | C7  |
|--------------|----|----|---------|----|----|----|-----|
| BBC          | 2  | 2  | S3 / T1 | 1  | 1  | 0  | No  |
| Reuters      | 2  | 4  | S3 / T1 | 1  | 2  | 0  | No  |
| The Guardian | 2  | 1  | S3 / T1 | 1  | 2  | 0  | No  |
| Mirror       | 2  | 2  | S3 / T1 | 1  | 2  | 0  | No  |
| Daily Mail   | 2  | 4  | S3 / T1 | 1  | 3  | 0  | Yes |
| The Sun      | 2  | 1  | S3 / T1 | 1  | 3  | 0  | Yes |

In terms of medical specificity, in all studied media outlets we can observe the disclosure of the treatment, framed as “preventative chemotherapy” without a clearer than before diagnosis. No stage or type of cancer is mentioned. All studied media outlets, except for the *Daily Mail*, use the same language as the language present in the written statement. *The Daily Mail* underlines the lack of specificity of the information offered, recasting the same tension that we have previously observed in its case between the ‘body politic’ and the ‘body natural’. When discerning between the media outlets’ approach to the reasoning behind non-disclosure, we observe that *Reuters*

and the *Daily Mail* explicitly mark the tension with the institutional framework, pointing out the fact that the palace “declined to give any details about the type of cancer”. The Princess is portrayed in her role as agent by the *Daily Mail*, which stresses, on multiple occasions, the fact that she has personally asked for “time, space and privacy”. *The Guardian* and *The Sun* align in offering the explanation on not placing additional pressure on the family. Similarly, the *BBC* and *Mirror* explicitly mention the scope of protecting the children. In terms of identifying the speaker and dominant identity type, all media outlets align once again in their representation of a mixed voice that includes the Princess’ first-person statement mixed with third-person institutional narration. The dominant identity type observed is the ‘body natural’, framed as patient, mother and wife. The ‘body politic’ remains present via the performative act of attending the public event, as the balcony, as well as her rank as Colonel of the Irish Guards are mentioned. We therefore observe directly, as the duality of the body natural and body political is simultaneously performed.

The common approach regarding advocacy or patronage is that of expressing gratitude for the citizens who have expressed solidarity in support, without transforming her illness into a formal cause. The shared however private experience is virtually expressed in order to be transformed into a form of affective capital. The fifth criteria pertaining to external pressure is linked to the degree in which each media outlet reiterates previous speculations regarding “where is Kate?”. *Reuters*, *The Guardian* and *Mirror* reiterate the “cancer scare” correlated with the lengthy absence. *The Guardian* underlines “the extra pressure” that the family was subjected to due to various speculations. Sitting at a more extreme end of the spectrum, *The Sun* and *Daily Mail* reiterated the audience’s speculations. There is an article published by *The Sun* conveying speculations almost in its entirety. We can conclude that the registered tension spans from editorial choice, rather than from the statement itself.

Performing the qualitative analysis of the selected sample from Stuart Hall’s reception and decoding perspectives, we observe that all six studied media outlets adopt a dominant position, conveying the statement’s initial intent of informing further the audiences regarding the measured progress. The oppositional decoding is performed minimally and contained in both *BBC*’s and *Reuters*’s case. *The Sun* actively reinforced the Palace’s own request for privacy, its decoding of the message reading still aligned with the original intent of the message.

The study of the media outlets' respect afforded to the right to privacy, verified through the prism of the criteria stated by the ECtHR in order to perform the balancing act between the two fundamental rights to privacy and family life and freedom of expression, resulted in the majority of outlets being classified as respecting the right to privacy. Registering an interesting yet predictable evolution, the *BBC*, *Reuters*, *The Guardian* and *Mirror* treated the health of the Princess as a matter of legitimate public interest. At the opposite end of the spectrum, *The Sun* admits “exclusively revealing the video of Kate and William”, thus proudly confessing the surveillance of private life and using the fable justification of a public interest. In this case we can observe a clear similarity with the Von Hannover 2 case, leading us into observing that, perhaps, the European Court should be offered also the possibility to establish a new threshold for the criteria of public interest to be adequately refined, in light of the transparency and consensual sharing of information that we have observed through this case study.

#### 4.4.2.5. Media framing of the announcement of the end of preventive chemotherapy - September 2024

A corpus comprising fourteen media products published by the six media outlets was selected in order to study the final reception of the body natural, as the news of the end of preventive chemotherapy is shared. The table below presents the aggregated results obtained by performing the qualitative analysis of the selected articles.

**Table 6.** Qualitative analysis

| Media Outlet        | C1 | C2  | C3 (Subtype/Type) | C4  | C5  | C6  | C7     |
|---------------------|----|-----|-------------------|-----|-----|-----|--------|
| <b>BBC News</b>     | 2  | 4   | S3 / T1-T2        | 1   | 2   | 0-1 | No     |
| <b>Reuters</b>      | 2  | 2-4 | S2-S3 / T1        | 1   | 1-2 | 0   | No     |
| <b>The Guardian</b> | 2  | 0-4 | S1-S3 / T1        | 0-1 | 1-2 | 0-1 | Yes/No |
| <b>The Sun</b>      | 2  | 2   | S3 / T1           | 1   | 2-3 | 0   | Yes    |
| <b>Mirror</b>       | 2  | 1-2 | S1-S3 / T1        | 0-1 | 1-2 | 0-1 | No/Yes |
| <b>Daily Mail</b>   | 2  | 4   | S2-S3 / T1        | 1   | 1-2 | 0-1 | Yes/No |

The analysis performed in order to discover the medical information disclosed and its level of specificity showed, repeatedly, for all studied media outlets, the same degree of conformity as previously observed. The treatment, framed as preventive, is stated within all studied media products included in the sample; however, the type and stage of the cancer remain undisclosed. We can conclude that the 'body politic' strictly controls also the limits of the 'body natural', allowing it to appear in a patient role, whilst preserving the strict non-disclosure standard regarding the diagnostic. The choice of justification regarding non-disclosure remains split between privacy ownership as observed by the *BBC* and *Reuters*, and the initial appeal to protection of children, as observed by *The Sun* and *Mirror*. The analysis of the speaker and dominant identity type reflected by the media coverage indicate that the majority of the studied media outlets conveyed a mixed institutional and subjective voice. The 'body natural' represented by patient, mother and wife is omnipresent.

The advocacy dimension recorded discursive solidarity across all studied media outlets. The disclosure tone of the articles is meant to enhance affective public capital, as we observe across all six studied media outlets. In connection with Sandra Petronio's borderline turbulence, we observe it at its lowest value when we analyze the texts that are closest to the original statement. Examples of these include *Reuter's* transcript and the *Daily Mail's* "message in full". Turbulence is recorded however at its highest when analyzing *The Sun's* article focusing on Prince William, which also is the only media product reassembling the entire history of the case. As previously observed, pressure relates to editorial memory rather than the overtly calm September announcement.

In terms of Stuart Hall's reception theory, the majority of the studied media products were marked as decoding dominantly. As evidence of this, *Reuter's* pieces, the *Mirror's* ones and the *Daily Mail* media products, they all decoded and conveyed to the audiences the same original institutional celebratory frame. Less than a third of the analyzed media products decoded in a negotiated frame, presenting clinical caveats, like, for example, in one in *The Guardian*.

The discussion regarding intrusions into privacy based on the Von Hannover criteria is no longer one of relevance. The clear positioning, through both bodies, natural and political, is that of sharing a limited consent for intrusion into privacy, as it is seen as a necessary sacrifice to reduce tension and to ensure that the dominant institutional message is decoded accordingly.

## 5. Conclusions

To conclude and answer the second research question in this study – related to the gradual complex and dynamic process of identity construction, there is evident change within the relationship private – institutional, in the case discussed here. The dichotomy follows a transformative process of construction and management, in four phases, from the initial ‘controlled institutional identity’, to ‘diffused identity’, respectively to the ‘assumed private identity’ and lastly to the ‘converted type of identity’ in which the private becomes an institutional capital.

The four theoretical concepts that have been operationalized across the current study, in order to discuss the data gathered in the seven official press releases about the Princess of Wales’s cancer experience, could be summed up as follows: Ernst Kantorowicz’ duality ‘body natural’ – ‘body political’ evidently defines the characteristics and problematic of the relationship person-function, with the function transcending the individual; Erving Goffman’s performative self sees the person as a creator of impressions for the public, where the self is not innately possessed but rather produced in the dialogue front stage – back stage, within the interaction between Catherine’s media messages and her public; Sandra Petronio’s *Boundaries of Privacy* and the Communication Privacy Management, outlined as a pattern to follow, have provided the relational/ communicational perspective of analysis, therefore allowing for an interpretation of the public’s access to the princess’ private realm. Stuart Hall’s approach, with his epistemological debate on identity as discourse, unites the theoretical and empirical data of the present study, seeing Catherine’s identity as strategically constructed and communicated while the ‘subject’ is ‘interpellated’ across and by certain media discourses.

With regard to RQ3 and RQ4 – that aimed at exploring how the selected media outlets received, decoded and further recoded and imparted the information – the sample selected for analysis can be clearly divided into two clusters of outlets. *The BBC*, *Reuters* and *The Guardian* provided their audiences with perspectives that can consistently be interpreted as dominant decoders, by conveying the Palace’s and the Princess’ intended framing. Therefore, the institutional identity, or ‘body political’, is seen as notifying the media and the citizens of the intended withdrawal from the performance of public duties. The negotiated positioning, rarely occurring, tends to criticize circumstantial rather than fundamental meaning, as it was the case of

expressing skepticism regarding the timing of the video confessing the vague cancer diagnosis. By contrast, *The Sun*, *Mirror* and *Daily Mail* presented, throughout the analysis, a significant element of unpredictability. Although initially subscribing to the dominant reception during the “abdominal surgery” and subsequent hospitalisation period, as well as when relaying the confession regarding the cancerous nature of the health problems, these media outlets shift towards the negotiated or even oppositional decoding during the edited Mother’s Day controversy. A potential conclusion can be drawn from these observations, namely the confirmation that when the factual basis illustrates an alignment between the political and the natural body, as well as the intended readability encoded in the initial message, these media outlets subscribe to the dominant frame. However, when the misalignment between the natural and the political body becomes apparent, their decoding shifts towards the negotiated and oppositional type of decoding of the message, in an attempt to discover reason and offer enriched meanings to the audiences, even if it relies on speculation, rather than facts. The separate legal dimension of the analysis of the selected media products, aimed at discovering whether we can clearly identify elements of the balancing criteria from the Von Hannover jurisprudence, proved not to be relevant for the study. Most of the media products illustrate a clear respect towards the institutionally and personally requested right to privacy, while sharing medical information that was consensually disclosed and while responding to legitimate public interest.

There are certain aspects that have not been viewed in this exploratory perspective: the paper does not discuss the authenticity of exposure in the corpus of analysis of the official press releases. There is no in-depth focus on who the ‘real’ author of the manipulated photo is, for example, or one may further debate on whether Catherine’s private story in the diagnosis disclosure video is authentic, or it is rather a performance of authenticity. It would also be interesting to see what happens if the conversion of private identity into a resource for institutional identity construction, through advocacy, might be reversible.

Several methodological and scope-related limitations of the qualitative media monitoring need to be acknowledged. While the selection of the analyzed media institutions and media products has been restrictively developed in terms of the surveyed dates, in order to include the date of specific official statements and events, the negotiated and oppositional decoding frames, in

Hall's sense, often require a longer period to fully develop, since reinterpretation, commentary, and counter-narratives tend to evolve over time. On a similar note, the corpus of analysis did not include articles built around the narrative perspectives of the extended royal family. Such an inclusion could have contributed to an improved analysis and understanding of the tension between the 'body natural' and the 'body politic'. The study is not complemented by an analysis of audience comments posted online. Considering Stuart Hall's reception theory, a fully rounded study should analyze not only the decoding and recoding of the original message performed by the media, but also the study of the decoding performed by the audiences of media. Future research incorporating audience-level data would therefore be necessary to complete the reception circuit and to test whether the decoding positions of the media, as identified here, correspond to the patterns of audience reception.

Lastly, for further research, it would be interesting to compare this specific case of identity construction and management regarding the duality 'body natural' – 'body political' with similar media phenomena in connection with the Princess of Wales and / or other British royal figures.

## References

- Deringer, Kathryn F., *Privacy and the Press: The Convergence of British and French Law in Accordance with the European Convention of Human Rights*, 22 Penn St. Int'l L. Rev. 191 (2003).  
<https://insight.dickinsonlaw.psu.edu/cgi/viewcontent.cgi?article=1615&context=psilr>
- Goffman, Erving. (1959). *The Presentation of the Self in Everyday Life*. Doubleday Anchor Books.
- Hall, Stuart. (2003). "Who Needs 'Identity'?" in Hall, Stuart and Paul Du Gay (eds.) (2003) *Questions of Cultural Identity*. Sage Publications.
- Kantorowicz, Ernst. (1957). *The King's Two Bodies. A Study in Mediaeval Political Theology*. Princeton University Press.
- Petronio, Sandra. (2002). *Boundaries of Privacy. Dialectics of Disclosure*. State University of New York Press.
- Rollo-Koster, J. (2017, October 2). body politic. *Encyclopedia Britannica*.  
<https://www.britannica.com/topic/body-politic>
- Wilson, Frances. (2024, March 27). "Kate's Two Bodies". *The New York Review*.  
[https://www.nybooks.com/online/2024/03/27/kate-middleton-two-bodies/?lp\\_txn\\_id=1682594](https://www.nybooks.com/online/2024/03/27/kate-middleton-two-bodies/?lp_txn_id=1682594)

### **ECHR Jurisprudence:**

Al Fayed v France, Application no. 38501/02, available at

<https://hudoc.echr.coe.int/eng?i=001-83547>

Von Hannover v Germany, Application no. 59320/00, available at

<https://hudoc.echr.coe.int/eng?i=001-61853>

Von Hannover v Germany, Application nos. 40660/08 and 60641/08, available at

<https://hudoc.echr.coe.int/eng?i=001-109029>

### **Official Press Releases:**

Prince and Princess of Wales. [@princeandprincessofwales]. Instagram. Available at  
<https://www.instagram.com/princeandprincessofwales/?hl=en#>

Prince and Princess of Wales. [@princeandprincessofwales]. (2024, January 17). *A Statement from Kensington Palace*. [official statement]. Instagram. Available at  
<https://www.instagram.com/p/C2NDoYrN-9r/?hl=en>

Prince and Princess of Wales. [@princeandprincessofwales]. (2024, January 29). [text] *A Statement from Kensington Palace*. [official statement] Instagram. Available at  
<https://www.instagram.com/p/C2rvObRtIO8/?hl=en>

Prince and Princess of Wales. [@princeandprincessofwales]. (2024, March 10). *Thank you for your kind wishes and continued support over the last two months. Wishing everyone a Happy Mother's Day. C*. [photograph and caption]. Instagram. Available at  
[https://www.instagram.com/p/C4U\\_IqTNaqU/?hl=en](https://www.instagram.com/p/C4U_IqTNaqU/?hl=en)

Prince and Princess of Wales. [@princeandprincessofwales]. (2024, March 22). *A message from Catherine, The Princess of Wales*. [video and subtitles]. Instagram. Available at  
<https://www.instagram.com/p/C402JKPtLVB/?hl=en>

Prince and Princess of Wales. [@princeandprincessofwales]. (2024, June 14). *I have been blown away by all the kind messages of support and encouragement [...]* [photograph and caption]. Instagram. Available at  
<https://www.instagram.com/p/C8NCRQKt4fA/?hl=en>

Prince and Princess of Wales. [@princeandprincessofwales]. (2024, September 9). *A message from Catherine, The Princess of Wales*. [video and caption]. Instagram. Available at  
[https://www.instagram.com/p/C\\_s6SukNNne/?hl=en](https://www.instagram.com/p/C_s6SukNNne/?hl=en)

Prince and Princess of Wales. [@princeandprincessofwales]. (2025, January 14). *I wanted to take the opportunity to say thank you to the Royal Marsden [...]*. [photograph and caption]. Instagram. Available at  
<https://www.instagram.com/p/DE0BZyJN1kt/?hl=en>

**Newsmedia Websites:**

[www.bbc.com/](http://www.bbc.com/)

[www.dailymail.com/](http://www.dailymail.com/)

[www.mirror.co.uk/](http://www.mirror.co.uk/)

[www.reuters.com/](http://www.reuters.com/)

[www.theguardian.com/](http://www.theguardian.com/)

[www.thesun.co.uk/](http://www.thesun.co.uk/)

## Deepfakes, Fake News, And Digital Trust: Insights from Romanian Social Media Users

Mirela ABRUDAN<sup>1</sup> , Carmen Ioana NEAMȚU<sup>2</sup>

**Abstract.** In today's digital society, the spread of misinformation through false online content has become a major problem worldwide. Fake news, deceiving information presented as real, has the potential to manipulate public opinion, distort facts, and weaken trust in traditional and online media sources. With technological advances, a new form of misinformation has emerged known as deepfakes, which makes distinguishing between reality and fiction even more challenging. Deepfakes are a more complex form of digital manipulation, using artificial intelligence to generate apparently genuine but in fact completely fake audiovisual content. The present study analyses the impact of fake news and deepfakes on Romanian social media users by conducting a non-representative self-administered online survey (N=154). The results indicate that the effects of fake news and deepfakes are influenced by socio-demographic factors such as age, gender, and educational level. Young people, as well as Romanian social media users with a high level of education, are more familiar and aware of these phenomena and their negative effects, likely due to higher media literacy. The collected data also suggest that female users have a critical and cautious approach regarding the consumption of digital content, thereby reducing the impact of misinformation. Romanians also believe that repeated exposure to fake news and deepfakes

---

<sup>1</sup> Assoc. Prof. Ph.D., Department of Communication, PR, and Advertising, Babeș-Bolyai University, mirela.abrudan@fspac.ro

<sup>2</sup> MA Communication and Advertising, National School of Political and Administrative Studies SNSPA, ioananeamtu41@yahoo.com.



leads to a loss of trust in social media. However, most users do not consider that being repeatedly confronted with fake news and deepfakes makes them necessarily more susceptible to manipulation.

**Keywords:** fake news, deepfakes, misinformation, disinformation, manipulation, social media, Romania.

## Introduction

In recent years, communication technologies and artificial intelligence have amplified the propagation of fake news and deepfakes in the digital environment. Due to the outbreak of the COVID-19 pandemic, the war started by Russia in Ukraine, but most of all the electoral campaign, followed by the annulment of the first round of the 2024 presidential elections, disinformation and manipulation of public opinion also reached alarming levels in Romania (Stănescu, 2024). Thus, the spread of false content poses a major challenge to the integrity of public discourse and the stability of democratic societies. Fake news, deceiving information presented as real, uses social media platforms to spread rapidly and often manipulates public opinion for political, financial, or ideological purposes. Deepfakes, a more advanced form of technology using artificial intelligence to produce fake video or audio content, further complicate the information landscape by undermining trust in visual and auditory evidence. Taken together, these forms of disinformation undermine public trust in media institutions and democratic processes, highlighting the urgent need for robust detection and awareness mechanisms to mitigate their impact (Botha & Pieterse, 2020). In October 2022, the EU adopted the Digital Services Act, which should have been fully effective by February 2024, a regulation applied to online intermediaries, including social networks, meant also to tackle disinformation. To do so, online platforms must hire fact-checkers, maintain content moderation teams across all official EU languages, and implement rapid alert systems, especially around elections. Content generated by artificial intelligence, such as deepfakes, must also be marked correspondingly (O'Carroll, 2024).

Numerous studies and reports have explored how fake news and deepfakes affect Romanian social media users. The most analysed topics in the context of digital communication and fake content rank among exposure and

vulnerability perceptions; engagement on online platforms generated by fake content; fake news, algorithms, and political influence; deepfake and cyberbullying; deepfake scams, fraud and propaganda, or election manipulation.

A national survey conducted by INSCOP Research at the request of the press agency News.ro (2024) found that over 70% of Romanians believe they are exposed to fake news, and 43% say social networks present the highest perceived risk for disinformation. Moreover, 75.9% of the Romanian citizens think that their voting choices are affected by disinformation and fake news (INSCOP, 2025). Reuters Institute Digital News Report from 2024 emphasises Romanians' reliance on online platforms when it comes to news consumption, and the rising concern about misinformation (Radu, 2024). The high occurrence of manipulation mechanisms on Romanian social media during the 2024 elections was marked as an unprecedented phenomenon in the country's history, resulting in the annulment of the first round of the presidential election, due to the electoral process being deemed as flawed. Online disinformation distorted public opinion and agenda-setting (Stănescu, 2024). TikTok was used for at-scale manipulation attempts to reach Romanian users (Global Witness, 2024). Network analysis shows bot-driven amplification on anti-EU / NATO narratives and cites around 34% fake-profile engagement on key posts, illustrating how coordinated inauthentic behaviour can shape perceived sentiment among social media audiences (Baruchin, 2025). Bitdefender Labs reported numerous scam lures during 2024, from fake celebrity endorsements to AI audio and video deepfakes, cloned media brands and fabricated financial testimonials used in social media ads, advising Romanians how to manage such practices (Bizgă, 2025). The 2025 Digital News Report highlights audiences' growing concern about synthetic content and misinformation, reinforcing why Romanian social media users report vigilance and avoidance (Newman, 2025).

As the spread of fake news and deepfakes continues to pose significant challenges to social trust, media integrity, and democratic processes, individuals, institutions, and governments must urgently develop critical thinking skills, media literacy initiatives, and technological solutions to combat the propagation of falsehoods and defend the reliability of public discourse.

On the 1<sup>st</sup> of April 2024, the Romanian National Cybersecurity Directorate released a guide designed to help identify deepfake content. The document provides useful information on the importance of understanding the phenomenon, technologies behind it, practical examples and warning signs,

detection tools and methods, the role of education in countering deepfakes, along with useful awareness tips, and recommendations in case of becoming a deepfake victim (JURIDICE.ro, 2024).

In this context, the current study surveys Romanian social media users to map their awareness and perceptions of negative influences of fake news and deepfakes in the digital environment.

## **Fake news**

Although the concept of fake news is often associated with the digital age and the emergence of online platforms, the spread of misinformation and propaganda precedes the Internet by many centuries (Winston & Winston, 2020). From ancient times, misinformation, disinformation, and propaganda have shaped public perception, influenced political discourse, and had a profound impact on social dynamics (Soll, 2016). The principle of fake news can be traced back to ancient civilisations, where several means of propaganda were used by rulers and elites to maintain power and control over the masses. Political leaders and religious authorities resorted throughout history to censorship, manipulation of historical narratives, and the dissemination of fabricated stories to strengthen their authority and suppress dissenting opinions (Soll, 2016).

Over the past decade, the concept of fake news has become a prominent and controversial topic on the media scene, as well as in politics and social discourse. Coined in the late 2000s and propagated during the turbulent political events of the 2016 presidential election in the U.S., fake news has evolved from a niche term to a ubiquitous and multifaceted phenomenon with far-reaching implications for public trust, the integrity of information, and democratic processes (Allcott & Gentzkow, 2017).

The propagation of misleading narratives and conspiracy theories across social media platforms, along with concerns about foreign interference and algorithmic amplification, influenced public opinion and generated a global debate about the spread and impact of misinformation in the digital age (Allcott & Gentzkow, 2017; Tandoc et al., 2018). In response, academics, journalists, policymakers, and technology companies have joined their efforts to limit and combat the dissemination of fake news through interdisciplinary research, fact-checking initiatives, media literacy programs, and regulatory intervention (Tandoc et al., 2018).

Despite these efforts, the phenomenon remains a challenge today, fuelled by partisan polarisation, echo chambers, and information silos, extending beyond politics to public health, climate change, social cohesion, and presenting complex ethical, legal, and technological challenges for actors in various fields (Gelfert, 2018). The study of fake news represents, therefore, a dynamic and interdisciplinary area of research, reflecting the urgent need to develop mechanisms for combating misinformation in an increasingly connected and digital world.

Fake news encompasses not only inaccuracies but also content deliberately created to manipulate and influence journalistic discourse (Rini, 2017). It may contain partially incorrect information due to journalistic negligence, whether because of the lack of experience, or failure to check sources. However, in such cases, journalists do not intend to mislead their audience or hide the truth (Baptista & Gradim, 2022). Although most scholars support the theory that fake news is not entirely fabricated stories, some voices support this view (Buckingham, 2019). Fake news uses sensational, provocative, and controversial narratives that aim to exploit social stereotypes and prejudices, thus producing strong emotional responses (Rochlin, 2017). This approach facilitates the virality of false information, as such content is more likely to be shared. The profitability of fake news depends on its popularity, with Rini (2017) arguing that it is a form of mass manipulation and explaining that the degree of untruthfulness is vital for its perpetuation, whether for ideological or pecuniary reasons.

The phenomenon of fake news is a complex and multifaceted challenge with profound implications for society, democracy, and information integrity. It represents more than just misinformation, involving the deliberate manipulation of facts, the dissemination of false representations, and the exploitation of societal vulnerabilities for various purposes, whether political, ideological, or financial (Gelfert, 2018). Furthermore, the dissemination of fake news in the digital age underscores the need for robust strategies to combat misinformation, including initiatives to improve media literacy, fact-checking mechanisms, and ethical journalistic practices. Critical thinking and digital competence are important for the public to distinguish facts from fiction in an increasingly complex media landscape. Tackling the problem of fake news requires joint efforts by various actors, including governments, technology companies, media organisations, and civil society, to promote a better-informed, more resilient, and more democratic society (Tandoc et al., 2018).

## **Fake news on social media**

In today's digital world, the propagation of fake news via social networking sites is a complex and multifaceted process. As individuals navigate a flood of information, the dissemination of misleading or fabricated content can cause confusion, disruption, and undermine trust in societal institutions (Arin et al., 2023). The high occurrence of false information on online platforms has concrete effects on individuals' beliefs and behaviour, and even on democratic processes. Creators and providers of such content can exploit vulnerabilities in social media algorithms to manipulate public discourse for ideological, political, or financial purposes (Arin et al., 2023).

The accessibility and affordability of creating social media profiles also contribute to the proliferation of harmful actors, like social bots, cyborg users, and trolls (Shu et al., 2017). Social bots, automated accounts controlled by computer algorithms, work autonomously to create content and interact with human users or other bots on social networks (Ferrara et al., 2016; Shu et al., 2017). These entities can play a harmful role by deliberately spreading misinformation and manipulating online discourse, as demonstrated by their widespread influence during the electoral campaign before the 2016 U.S. presidential elections. The decentralised nature of Twitter, officially known as X since 2023, and its rapid transmission capabilities, make it vulnerable to the spread of misinformation and disinformation (Drolsbach & Pröllochs, 2023). Fake news also has the potential to destabilise the integrity of the news ecosystem by eroding its foundations of authenticity and trust. For example, during the 2016 U.S. presidential elections, it became clear that the most popular deceiving stories on Facebook were shared more than the most popular authentic news in the mainstream media (Allcott & Gentzkow, 2017).

At the same time, trolls, real human users who are experts at causing chaos, disrupting online communities and provoking emotional reactions from users, also exert a significant influence on the propagation of fake news in the digital realm. Their deliberate intention to arouse negative emotions such as anger and fear, which in turn cause doubt, mistrust, and irrational behaviour among users (Shu et al., 2017), exacerbates the dissemination of disinformation in digital spaces and highlights the complex interplay between human behaviour and technological platforms in today's information ecosystems (Cheng et al., 2017). According to reports, there were around 1,000 paid Russian trolls involved in spreading fake news about Hillary Clinton. The behaviour

of trolls is particularly influenced by the emotional state of users and the contextual dynamics of online conversations, which facilitates the smooth spread of false stories in seemingly “natural” digital communities (Shu et al., 2017).

Cyborg accounts are created by humans as a cover, with automated programs performing actions on online platforms (Ferrara et al., 2016). The fluid transferability between human and automated functionality offers cyborg users various opportunities to spread disinformation. Overall, these actively partisan and productive malicious accounts exert considerable influence as powerful and amplifying sources of fake news in social media ecosystems (Shu et al., 2017).

With the growing importance of social networks, people have become increasingly reliant on these digital channels to keep up to date with current events. However, a major problem permeating these news feeds is the ubiquitous presence of false or misleading content. Despite the efforts of some online platforms to curb the propagation of fake news through features such as tagging or warnings, the persistence of false information has profound and lasting effects. Fake news not only misleads people but also creates a climate of uncertainty and mistrust that undermines the very foundations of informed public discourse. Furthermore, its uncontrolled spread can fuel social unrest, exacerbate existing tensions, and even hinder democratic processes by undermining the integrity of public discourse and decision-making in elections (Arin et al., 2023).

## **Deepfake**

The emergence and propagation of deepfakes have generated both excitement and concern in the fields of technology, media, politics and society in general. Deepfakes are a sophisticated form of synthetic media produced through artificial intelligence-based algorithms, whereby existing images or videos are digitally manipulated or overlaid in such a way that highly convincing but completely fake content is created. This innovative technology blurs the lines between facts and fiction and has profound ethical, social, and political implications (Narayan et al., 2022). While deepfakes offer new opportunities for creative expression and entertainment, their potential for abuse and exploitation raises critical questions about misinformation, privacy

violations, and the loss of trust in both visual and audio media. Therefore, understanding the complexity and implications of this phenomenon is important for policymakers, media professionals, and the public (Chesney & Citron, 2019).

Deepfakes convincingly imitate human behaviour and appearance, extending from visual environments to audio manipulation. Technological progress has made deepfakes almost indistinguishable from real images, presenting a range of challenges and opportunities for people around the world. The term “deepfake” originates from the fusion of “deep learning” and “fake” and illustrates the dependence on artificial neural networks in machine learning (Mirsky & Lee, 2021). Primarily used to manipulate and generate human images, this technology also has a variety of creative and useful functions. It can be found in numerous sectors, such as the fashion industry and e-commerce, where it simplifies consumer transactions. In addition, the entertainment industry uses deepfake technology to synthesise artificial voices for musicians and enhance film productions with sophisticated special effects (Chesney & Citron, 2019). Its therapeutic potential is also evident, considering the ability to enable simulated interactions for Alzheimer patients, potentially helping to preserve memory (Mankoo, 2023). In addition, various online communities have emerged that explore the comedic potential of deepfake memes, such as the creation of music videos.

Despite the positive outlook, however, deepfake technology has also become notorious for its unethical and harmful aspects. In November 2017, the first widely known case of deepfakes appeared, when a Reddit user known as “Deepfakes” used deep learning algorithms to insert the faces of celebrities into pornographic videos, sparking a media frenzy and an increase in the creation of similar content (Mirsky & Lee, 2021). Subsequent cases, such as the spread of a deepfake video file showing former U.S. president Barack Obama, have intensified fears about identity theft and the propagation of falsehoods on online platforms. Starting with these incidents, academic interest in deepfakes has grown exponentially (Barari et al., 2021), due to the significant ethical dilemmas and societal risks that the phenomenon presents. Prominent figures such as politicians and celebrities are particularly vulnerable to the creation of defamatory or pornographic deepfake content, which exploits the publicly available media in which they are portrayed. This ethical issue is exacerbated by the ease with which individuals can create misleading content, which can lead to malicious abuse, including the dissemination of

unauthorised or fake images and audio files. Consequently, the spread of untraceable artificial content poses a major challenge to authenticity, integrity, and privacy in the digital environment (Akhtar, 2023).

## **Deepfake on social media**

Even if technological advancement has generated new possibilities for creative expression, entertainment, and educational purposes, the propagation of deepfakes on social networking sites poses formidable challenges, including the potential for widespread misrepresentation, manipulation of public perception, and undermining trust in digital media (Chesney & Citron, 2019). Furthermore, deepfakes raise profound ethical and societal questions, such as privacy violations, identity theft, and confusion between truth and fiction. Given the evolution and propagation of deepfake technology, researchers are working to comprehensively understand its multifaceted repercussions on the dynamics of social media, the integrity of information, and broader societal discourse. Through rigorous research and analysis, scientists are attempting to shed light on the complexity of deepfake technology and develop strategies that aim to diminish its negative consequences while harnessing its capacity for positive societal impact (Akhtar, 2023).

The dissemination of digital photos and videos has increased significantly due to the widespread use of smartphones and the growing popularity of online platforms. This has led to social networks becoming one of the most important sources of daily information and news consumption. Among these platforms, X, formerly Twitter, has a decentralised and dynamic nature and stands out with a significant active user base, particularly as a fast news disseminator compared to traditional media (Fagni et al., 2021). However, this also leads to an increase in spoofing techniques aimed at quickly spreading false information. The fake content generated is carefully designed to appear authentic and often focuses on current topics to accelerate its spread and involve more people in false stories, ultimately posing significant risks to society (Tolosana et al., 2020). Such multimedia content can extend widely in an uncontrolled manner and creates an alternative narrative of reality, while identifying the originators of rumours remains a major challenge due to the difficulty of determining the source of the published media material (Narayan et al., 2022).

Also, on Facebook, one of the world's most popular social media channels, there is a serious risk that deepfakes could be used to propagate inaccurate information and distortions. This development not only calls into question the credibility of content but can also have a significant impact on user trust in the platform. Given these challenges, Facebook must develop effective strategies and technologies to detect and combat deepfakes while strengthening the awareness and digital media literacy of its users (Shu et al., 2017).

Fake content poses a significant digital threat. In times of heightened sensitivity or during major events, the propagation of falsehoods becomes more frequent. During elections, for example, there is an increased risk that opponents would exchange deepfake videos to undermine each other. Although deepfakes can have a profound impact on society, society itself plays a crucial role in combating this type of content (Gunther et al., 2018). The continued dissemination of fake videos on the Internet could undermine trust in online information for future generations. Society must therefore recognise that sharing deepfakes inadvertently helps their creators to spread misinformation. The social impact of deepfakes technology is profound and could lead to people being reluctant to distribute personal content on social networking sites due to trust issues, which could have far-reaching consequences for all industries (Chesney & Citron, 2019).

Given their immense influence, online platforms such as Facebook, Instagram, TikTok, or X have been viewed critically for their potential to quickly spread inappropriate content worldwide. Social networks have been used to promote terrorist violence, distort election results, and influence public perception of important democratic events.

Although working to implement strategies and technologies to combat fake news and deepfakes, the key challenge remains to ensure the authenticity of content and strengthen user trust in these platforms (Fagni et al., 2021). A comprehensive approach is needed, based on technological innovation, community collaboration, improved digital media literacy, and policies to regulate user engagement, to minimise the impact of disinformation in the digital realm.

## Research design

This research aims to identify the perceived effects that fake news and deepfakes occurring on online platforms might have on Romanian social media users from different age cohorts, gender, and educational backgrounds. To investigate whether individuals are aware of these phenomena and their possible negative impact, as well as the strategies they use to avoid misinformation and manipulation in the digital environment, a survey was designed and conducted.

Based on the literature at the intersection of social media use, effects of this behaviour, and fake news and deepfakes, a series of research questions and hypotheses were established, to be addressed via the quantitative research method employed:

**RQ1: Do socio-demographic criteria influence the recognition potential and the vigilance of Romanian social media users towards fake online content?**

*H1: The awareness and vigilance regarding fake news and deepfakes in the digital landscape are influenced by factors such as age, gender and, educational level.*

**RQ2: How are Romanian social media users affected by the exposure to fake news and deepfakes on social media?**

*H2: Users who are repeatedly exposed to fake news and deepfakes on social media are more susceptible to misinformation and manipulation.*

*H3: Frequent exposure to fake news and deepfakes on social media leads to a loss of public trust in these platforms and increases concern towards the consumption of manipulative online content.*

## Results

Through empirical analysis and critical examination, the current research intends to analyse the impact of fake news and deepfakes on Romanian social media users and thus, to contribute to a deeper understanding of the complex dynamics shaping today's digital communication and the consequences for individuals, communities, and democratic societies in general. Ultimately, the study aims to provide insights that can inform strategies to diminish the negative effects of fake news and deepfakes, to promote an informed and resilient information landscape in the digital age.

To address the research questions and hypotheses above, a quantitative research method was employed. A Romanian-language survey was developed and disseminated through Facebook and Instagram over the period April – May 2024. The questionnaire gathered N=154 valid responses. The sample was a non-probability one, constituted through self-selection, and the survey started with a section for participants' informed consent.

Socio-demographic data of the participants play an important role in the present research, given the formulated research questions and hypotheses. Regarding the age segmentation of the respondents, 3.25% fell into the age category of teenagers under 18. Young adults (18-24 years old) embody the most strongly represented age cohort in the sample (34.42%). Due to their high digital literacy and online media consumption habits, they probably easily accessed the questionnaire. 16.23% of the sample was represented by individuals aged 25 to 34, participants between 35 and 44 years old gathered 27.92%, while the age cohort between 45 and 60 also represented 16.23% of the total number of respondents. Finally, individuals over 60 years old are the least represented (1.95%), probably due to restricted access to the Internet or social media networks, a lower media literacy or interest in such online activities. They are also the only age category with no respondents aware of the concept of fake news or deepfakes. Young generations, under 18 to 44 years old, are instead familiar with the phenomena (93%), and indicate more vigilance when accessing information, especially on social networks, due to the emergence and spread of such content (87%).

In terms of gender representation, the collected data reveal a significant discrepancy, with most of the participants being females (66.2%). The male gender was less represented, with 52 respondents (33.8%). There were no registered responses of individuals declaring themselves non-binary. This strong presence of women could be due to their increased interest in the topic. The results also imply that women indicate a greater ability to recognise and reject false or inaccurate content (84%) than men (72%). A plausible hypothesis could be that females are more attentive to the phenomena of fake news and deepfakes, being more concerned about the accuracy and authenticity of online information.

Concerning the educational level, 29.2% of the survey participants declared pre-university studies as the last or current form of education attended. The majority have a bachelor's degree (32.5%) and a master's degree (24.7%), 8.4% of the respondents indicate having also postgraduate studies, while 5.2%

avoided answering this question. These numbers, corroborated with the data regarding the recognition potential and vigilance towards fake news and deepfakes in the online environment, could indicate a possible link between educational background and awareness of or interest in misinformation and manipulation in social media. Participants with higher educational levels, bachelor's or master's degree and post-university studies, are more familiar and recognise fake news and deepfakes on online platforms more easily (92%) than participants with pre-university studies (61%); this might also be because they are probably more familiar with the concepts of critical thinking and information evaluation. However, it is important to emphasise that this analysis provides only preliminary evidence and that we cannot draw definitive conclusions about the causal relationship between age, gender, and educational attainment and attitudes toward misinformation. It is plausible that other individual factors, such as personal experiences or sociocultural background, play an important role in influencing these attitudes.

The analysis of the frequency of social media use reveals a clear prevalence of users who access social media intensively and daily (76%). 15.6% of all respondents reported using social media daily, but not intensively. Only 8.4% of the participants use online social media platforms occasionally. However, the level of trust in content consumed on social networks is rather low. The data indicates a small proportion of the sample (1.9%) trusting the information on social media to a very high degree, and 11.7% to a high degree. 81.2% mentioned partial trust, and 4.5% of the respondents have no trust at all in the information posted on online social networking platforms. Collected data also shows that users are generally concerned about the truthfulness and reliability of online content (91,6%), and the analysis indicated that intensive use of social media is doubled by a stronger desire to be aware of the existence of misinformation and to understand how it can influence perceptions and decision-making (83,1%).

Regarding the familiarity with the concept of fake news, the survey participants (97.4%) are aware of and generally recognise this phenomenon and its negative effects in the online environment. A very small percentage (2.6%) stated that they are not familiar with this specific term. However, this does not necessarily imply that they are unaware of the phenomenon of misinformation itself. When asked about the frequency of encountering fake news on social media, 64.9% of respondents are confronted daily with this phenomenon, while 24.7% come across fake news several times a week.

Only 4.5% of the questioned social media users say they rarely encounter fake news on social networks. Knowledge of the concept of fake news can lead to a more vigilant attitude and a better ability to recognise and evaluate false or manipulative information on social media platforms. This correlation suggests that education about misinformation can play an important role in combating this phenomenon and promoting a more responsible and well-informed online experience.

Switching to deepfakes, 83.1% of the respondents stated they know the meaning of the concept and 54.5% consciously recognised such content on social media. This high level of awareness can be attributed to the growing interest in this technology among the media and the public. On the other hand, 11% of the survey participants are unaware of the existence of deepfakes, and 37% mentioned they have not consciously recognised a deepfake on social networking platforms so far. Finally, 5.8% of the respondents preferred not to answer the question referring to the familiarity of the concept of deepfakes, and 8.4% avoided indicating their recognition potential of such content. These findings underscore the need for a critical approach and adequate public education in detecting and dealing with digitally faked content and highlight the complexity and diversity of techniques used in the production of deepfakes.

Asked about how they deal with the discovery of fake digital content, the most common reaction of the survey participants was that they report the video to the platform where it was found (37.7%). This suggests that online platforms should be responsible for removing fake content and protecting users from its potential negative consequences. In addition, 29.9% of the respondents declared being cautious and checking the information using multiple sources before taking further action. This approach reflects a desire to confirm the authenticity of online content and avoid the dissemination of false information over the Internet. Data analysis indicates that the sources survey participants use and trust most for the validation process of information are official government sources (55.2%), followed by independent news sources (34.4%), mainstream media (31.8%) and personal relationships, like family and friends (24%). Online social networks are perceived as the least trustworthy source of information (14.3%). However, there is also a notable percentage (13.9%) of individuals in the sample who mentioned sharing and commenting on fake news and deepfakes immediately. Such a reaction could indicate a tendency to distribute information without first checking

its authenticity, which could lead to the unintentional rapid spread of fake content throughout the digital environment. Finally, 14.9% of the respondents simply ignore the discovery of fake news or deepfakes. The reasons for this kind of reaction could be manifold, including a lack of interest in or awareness of the impact and importance of the authenticity of online content. Overall, these findings emphasise once more the importance of education and public awareness about how to deal with and combat the propagation of fake news and deepfakes on the Internet.

Regarding the impact of false online content, 48.1% of the survey participants do not consider that the occurrence and the intensive exposure to fake news and deepfakes on social networks affect them in any way. This may reflect a certain confidence in their ability to evaluate and filter information, on the one hand. However, survey data indicate that only 9.7% of individuals questioned found it very easy to make the distinction between fake content and real content online. For almost half of the respondents (49.4%), the distinction is perceived as easy, but they acknowledge that it requires conscious effort and attention. On the contrary, around 32.5% of the participants find it difficult to identify the truthfulness of online information, while 3.2% consider this distinction to be very difficult. On the other hand, one explanation for the results concerning the effects of fake news and deepfakes on Romanian social media users can rely on the “third-person effect” (Lev-On, 2017), as most respondents (83.1%) declared themselves fully aware of the negative impact of misinformation on the society and find it very important to combat this phenomenon through information and education. Despite this statement, most individuals (48.1%) who expressed themselves regarding the influence on their own person denied registering any negative effects. On the contrary, 38.3% of the respondents admitted that frequent exposure to fake news and deepfakes makes them more susceptible to misinformation and manipulation. This significant percentage indicates that in Romanian online communities, there exists awareness of the potential negative effects of false information and digital manipulation on individuals’ thinking and behaviour. Fake digital content can also undermine trust in truthful information and distort the perception of reality. There is also a rather high number of individuals in the sample (13.6%) who avoided expressing their opinion concerning this important aspect of our survey.

The necessity of social media platforms taking measures to detect fake news and deepfakes is considered by 81.8% of the survey participants

as very important. In addition, 16.9% of the respondents recognised this issue as being important. There were no responses that described this subject as less important or not important at all. These results show a broad consensus among respondents on a topic intensely debated in the literature, namely the importance of preventive measures and the implementation of effective strategies and technologies for managing and filtering manipulative content on the Internet (Fagni et al., 2021).

Automatic detection technologies are considered an important solution for limiting the dissemination of misinformation on social networking sites by most respondents (67.5%). Almost half of the individuals questioned (46.1%) mentioned the importance of greater transparency regarding the functioning of online platforms and their algorithms for recommending and filtering content. Educating users to recognise and handle false information is also strongly supported by 35.1% of the survey participants, while strict sanctions for misinformation are considered important by 50% of them. Overall, the responses indicate that combating misinformation on social networks requires a combination of advanced technological solutions, transparency on the part of social media platforms, education and awareness-raising campaigns among users, and the introduction of stricter penalties for those who propagate false information.

The answers concerning the responsibility for the spread of fake news and deepfakes on the Internet reveal different views. Most respondents (71.4%) hold social media platforms accountable for the propagation of false content, while 51.3% believe that individuals and companies bear a large responsibility for the spread of misinformation in the digital environment. As for the government, 35.7% of the survey participants believe that it should bear some responsibility for combating the dissemination of misinformation on social networking platforms. This view reflects a desire for government intervention and regulation in online communication to protect users and promote a safer and more reliable digital environment.

Important negative effects of the analysed phenomena are also growing concerns and a significant loss of trust in the online environment due to the spread of false content. 19.5% of the respondents stated that their trust in social media has decreased significantly, and their concern regarding manipulation amplified due to frequent emergence and exposure to fake news and deepfakes, 31.2% lost trust to a great extent, while 39.6% of the individuals questioned declared that their trust in social networks has

decreased to some extent. These findings suggest that the credibility of online platforms is rather low among Romanian social media users and urge for rapid development of strategies and measures to reduce the negative effects of misinformation, manipulation, and propaganda on the Internet.

The last question deals with measures that could help restore trust in information on social media, and the responses correspond to the results obtained regarding the solutions for limiting the spread of misinformation and digital manipulation, which survey participants mentioned. The collected data indicate that, according to Romanian social media users, stricter verification of information (85.7%), user education (47.4%), and greater transparency of algorithms (29.2%) are the most important ways to restore people's trust in online platforms. These findings emphasise the urgent need for developing strategies and initiatives to improve the quality and reliability of online content, as well as an educational approach focused on raising awareness of the characteristics of false information and how to deal with it.

## **Discussion and conclusions**

The present research aimed to investigate the awareness and impact of fake news and deepfakes on Romanian social media users in the digital environment. If socio-demographic criteria influence the recognition potential and vigilance towards false and manipulative online content, or how intensive exposure to such content makes Romanian users more susceptible to misinformation and affects their trust in social networking sites, are some of the questions our endeavour tried to answer through an online self-administered survey. Despite the reliance on a convenience sample of limited size, which precludes a conclusive assessment of the hypotheses, the study nevertheless generates results of sufficient relevance to merit further discussion. Regarding RQ1 and the related hypothesis, the findings indicated that indeed variables like age, gender and educational background affect the awareness towards fake news and deepfakes in the digital environment. Thus, young generations, who grew up with technological progress and intensive use of social media platforms, have better skills to recognise these phenomena. Digital natives are generally more sceptical and vigilant about the information they consume online and use various methods and digital tools to verify its authenticity. Furthermore, Romanian female social media users seem to be aware of the risks associated with digital misinformation

to a greater extent than men, which prompts them to be cautious and reluctant to accept or share suspicious content. The level of education also plays an important role concerning awareness and understanding of fake news and deepfakes on the Internet. People with higher education are better informed, more familiar with the concepts, due to their higher media and digital literacy, and more likely to be able to recognise disinformation and complex digital manipulation.

H2 concerning the higher susceptibility to misinformation and manipulation in case of frequent exposure to fake news and deepfakes on social networks could not be confirmed, as most of the survey participants denied such influence. In this case, one of the factors that could have affected the validity of the results obtained is the phenomenon known as the “third-person effect hypothesis” (Bawazir & Idid, 2024). According to this, individuals tend to believe that they are less vulnerable to the influence of media messages than other people, and implicitly to an underestimation of the impact of fake news and deepfakes on oneself, while others are perceived as more susceptible. Thus, the questionnaire results are contradictory, because initially, most of the respondents acknowledged mistrust in social media, fuelled by frequent misinformation and sophisticated digital manipulation that undermines the credibility of information available online. Therefore, the collected data may not accurately reflect the actual perceptions and behaviours of users. This limitation could affect our understanding of the actual impact of disinformation phenomena in the Romanian digital landscape. However, acknowledging this may pave the way for further, more targeted, and longer-term research that more accurately examines the psychological effects of fake news and deepfakes on Internet users.

Romanian social media users are increasingly aware of the advanced techniques used to create fake news and especially deepfakes, which contribute to a widespread sense of distrust and caution. This rather general scepticism indicates that there is an urgent need for media education and the development of effective mechanisms for verifying information to restore trust in social media. Educational institutions, social media platforms, and regulatory authorities must work together to develop and implement media and digital literacy programs that equip users with the ability to recognise and combat misinformation. The results of the questionnaire analysis confirm, therefore, H3, that indeed, intensive exposure to fake news and deepfakes on social networking sites leads to a loss of public trust in these platforms and increases concern towards the consumption of manipulative online content.

As any such academic undertaking, the current study is inevitably constrained by certain methodological limitations and potential sources of error; nonetheless, it also elucidates promising directions for future research. Although the survey relied on a non-probabilistic, non-representative sample, the findings provide valuable insights that allow for a preliminary delineation of Romanian social media users' perception of fake news and deepfakes that occur daily in today's digital society. However, it must be mentioned that our study lacks a detailed analysis of the ethical perspective of the discussed topic, which is of great importance to understand the complexity of such phenomena. An ethical assessment would contribute significantly to clarifying the moral implications of the spread of disinformation and provide a solid basis for formulating counterstrategies that consider the principles of integrity and social responsibility. The advancement of deepfake technology can influence our perception of history and reality in general and poses major challenges to the understanding and interpretation of the past and present. The ability to create fake audio and video content with increasing accuracy raises fundamental questions about the authenticity and reliability of the information we consume, questions which are crucial for future research and critical engagement with the impact of deepfake technology in our society. Therefore, the ethical implications of the phenomena of fake news and deepfakes are profound and multifaceted, ranging from eroding public trust, manipulating political outcomes, and violating individuals' dignity and consent, to exposing regulatory inadequacies, endangering legal processes, and challenging norms around creativity and media literacy. Each dimension demands thoughtful solutions: legal frameworks, platform accountability, public education, and an ethical culture that balances innovation with responsibility.

## References

- Akhtar, Z. (2023). Deepfakes generation and detection: A short survey. *Journal of Imaging*, 9(1), 18. <https://doi.org/10.3390/jimaging9010018>
- Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211-236. DOI: 10.1257/jep.31.2.211
- Arin, K. P., Mazrekaj, D., & Thum, M. (2023). Ability of detecting and willingness to share fake news. *Scientific Reports*, 13(1), 7298. <https://doi.org/10.1038/s41598-023-34402-6>

- Baptista, J. P., & Gradim, A. (2022). A working definition of fake news. *Encyclopedia*, 2(1), 632–645. <https://doi.org/10.3390/encyclopedia2010043>
- Barari, S., Lucas, C., & Munger, K. (2021). Political Deepfake Videos Misinform the Public, But No More than Other Fake Media and (Sometimes) Real Media. *OSF Preprints*, 13, 1-16. <https://osf.io/cdfh3/>
- Baruchin, R. (2025, April 22). Analyzing Fake Profiles' Impact on Romania's Election Crisis. *Cyabra.com*. Retrieved from <https://cyabra.com/blog/analyzing-fake-profiles-impact-on-romania-s-election-crisis/>
- Bawazir, A. & Idid S. A. (2024). The impact of third-person effect on social networking sites privacy risks and protective measures adoption among Yemeni students in Malaysia. *Heliyon*, 10(9). DOI: 10.1016/j.heliyon.2024.e30386
- Bîzgă, A. (2025, May 16). Deepfake Scam Exploits Romanian Presidential Candidates to Lure Victims into Fake “Neptun Deep” Investment. *Bitdefender.com*. Retrieved from <https://www.bitdefender.com/en-us/blog/hotforsecurity/deepfake-scam-exploits-romanian-presidential-candidates-to-lure-victims-into-fake-neptun-deep-investment>
- Botha, J., & Pieterse, H. (2020). Fake News and Deepfakes: A Dangerous Threat for 21st Century Information Security. *ICCWS 2020 15<sup>th</sup> International Conference on Cyber Warfare and Security*, 57-66. DOI: 10.34190/ICCWS.20.085
- Buckingham, D. (2019). Teaching media in a ‘post-truth’ age: fake news, media bias and the challenge for media / digital literacy education / La enseñanza mediática en la era de la posverdad: fake news, sesgo mediático y el reto para la educación en materia de alfabetización mediática y digital. *Culture and Education/Cultura y Educación*, 31(2), 213–231. <https://doi.org/10.1080/11356405.2019.1603814>
- Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017). Anyone can become a troll: Causes of trolling behavior in online discussions. In Lee, C. P., & S. Poltrok (Eds.), *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing* (pp. 1217-1230). Association for Computer Machinery. <https://doi.org/10.1145/2998181.2998213>
- Chesney, R., & Citron, D. (2019). Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics. *Foreign Affairs*, 98, 147.
- Drolsbach, C. P., & Pröllochs, N. (2023). Diffusion of community fact-checked misinformation on Twitter. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2), 1–22. <https://doi.org/10.1145/3610058>
- Fagni, T., Falchi, F., Gambini, M., Martella, A., & Tesconi, M. (2021). TweepFake: About detecting deepfake tweets. *PLoS ONE*, 16(5), e0251415. DOI: 10.1371/journal.pone.0251415
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communication of the ACM*, 59(7), 96-104. <https://doi.org/10.1145/2818717>

- Gelfert, A. (2018). Fake news: A definition. *Informal Logic*, 38(1), 84-117.  
<https://doi.org/10.22329/il.v38i1.5068>
- Global Witness. (2024, December 17). What happened on TikTok around the annulled Romanian presidential election? An investigation and poll. *Author*. Retrieved from <https://globalwitness.org/en/campaigns/digital-threats/what-happened-on-tiktok-around-the-annulled-romanian-presidential-election-an-investigation-and-poll/>
- Gunther, R., Beck, P. A., & Nisbet, E. C. (2018). *Fake News May Have Contributed to Trump's 2016 Victory*. Ohio State University.  
<https://assets.documentcloud.org/documents/4429952/Fake-News-May-Have-Contributed-to-Trump-s-2016.pdf>
- INSCOP. (2025). January 2025 – INSCOP Research opinion poll, commissioned by Funky Citizens. Romania between national and European in the era of disinformation. Economic patriotism, values and conspiracies. *Author*. Retrieved from <https://www.inscop.ro/en/january-2025-inscop-research-opinion-poll-commissioned-by-funky-citizens/>
- JURIDICE.ro. (2024, April 2). Ghid privind identificarea materialelor de tip deepfake. Retrieved from <https://www.juridice.ro/731663/ghid-privind-identificarea-materialelor-de-tip-deepfake.html>
- Lev-On, A. (2017). The third-person effect on Facebook: The significance of perceived proficiency. *Telematics and Informatics*, 34(4), 252-260.  
<https://doi.org/10.1016/j.tele.2016.07.002>
- Mankoo, S. S. (2023). DeepFakes-The Digital Threat in the Real World. *Gyan Management Journal*, 17(1), 71-77. <https://doi.org/10.48165/gmj.2022.17.1.8>
- Mirsky, Y., & Lee, W. (2021). The Creation and Detection of Deepfakes: A Survey. *ACM Computing Surveys*, 54(1), 1-41. <https://doi.org/10.1145/3425780>
- Narayan, K., Agarwal, H., Mittal, S., Thakral, K., Kundu, S., Vatsa, M., & Singh, R. (2022). DeSI: Deepfake Source Identifier for Social Media. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (pp. 2858-2867).
- Newman, N. (2025). Overview and key findings of the Digital News Report. *Reuters Institute*. Retrieved from <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025/dnr-executive-summary>
- News.ro. (2024, July 10). Sondaj INSCOP la comanda News.ro – Mai mult de 70% dintre români consideră că sunt expuși știrilor false/ Peste 48% cred că în lume se întâmplă multe lucruri foarte importante, despre care oamenii obișnuiți nu sunt niciodată informați. *Author*. Retrieved from <https://www.news.ro/cultura-media/sondaj-inscop-comanda-news-mult-70-dintre-romani-considera-expusi-stirilor-false-pest-48-cred-lume-intampla-multe-lucruri-foarte-importante-oamenii-obisnuiti-niciodata-informati-1922400010002024070821661752>

- O'Carroll, L. (2024, March 26). Hire factcheckers to fight election fake news, EU tells tech firms. *The Guardian*. Retrieved from <https://www.theguardian.com/media/2024/mar/26/tech-firms-poised-to-mass-hire-factcheckers-before-eu-elections?>
- Radu, R.-N. (2024, June 17). Digital News Report. Romania. *Reuters Institute*. Retrieved from <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024/romania>
- Rini, R. (2017). Fake News and Partisan Epistemology. *Kennedy Institute of Ethics Journal*, 27(2), E-43-E-64. <https://doi.org/10.1353/ken.2017.0025>
- Rochlin, N. (2017). Fake news: belief in post-truth. *Library Hi Tech*, 35(3), 386-392. <https://doi.org/10.1108/lht-03-2017-0062>
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36. <https://doi.org/10.1145/3137597.3137600>
- Soll, J. (2016, December 18). The Long and Brutal History of Fake News. *Politico Magazine*. Retrieved from <https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535/>
- Stănescu, G. C. (2024). Fake News, Bots, and Influencers: The Impact of Social Media on Romania's 2024 Elections. *Social Sciences and Education Research Review*, 11(2), 361-366. <https://doi.org/10.5281/zenodo.15258337>
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "Fake News": A typology of scholarly definitions. *Digital Journalism*, 6(2), 137-153. <https://doi.org/10.1080/21670811.2017.1360143>
- Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A Survey of face manipulation and fake detection. *Information Fusion*, 64, 131-148. <https://doi.org/10.1016/j.inffus.2020.06.014>
- Winston, B., & Winston, M. (2020). *The Roots of Fake News. Objecting to Objective Journalism* (1<sup>st</sup> ed.). Routledge, London. <https://doi.org/10.4324/9780429032264>

# The Dual Construction of Generative Artificial Intelligence Domestication Practices and Information Cocoons. An Empirical Study Based on College Students' Usage Patterns

Liu QINGTAO<sup>1</sup>

**Abstract.** Does Generative AI (GenAI) reinforce or dismantle information cocoons? Moving beyond technological determinism, this study utilizes Domestication Theory to examine how user appropriation influences cognitive outcomes. A survey of 245 university students analyzed the effects of three GenAI use patterns – Instrumental, Information Acquisition, and Exploratory – mediated by AI Literacy and Critical Thinking. Results indicate that impact is practice-dependent: Information Acquisition Use intensifies cocoons (*cocoon-weaving*), while Exploratory Use diminishes them (*cocoon-breaking*). Notably, a literacy paradox emerges: high individual literacies paradoxically amplify the cocoon-weaving effect of information acquisition. The study confirms that use patterns influence cognitive outcomes both directly and through complex interactions with literacy. These findings extend domestication theory into the cognitive domain and suggest that educators must foster exploratory user practices to maintain cognitive openness in the AI era.

**Keywords:** Generative AI; information cocoons; domestication theory; AI literacy; media effects

---

<sup>1</sup> Lecturer, Hebei University of Economics and Business, liuqingtao80@126.com



## 1. Introduction

### *1.1 Research Background*

The phenomenon of information cocoons—characterized by information narrowing and opinion polarization—has emerged as a central concern in communication research and digital governance (Zhang, 2025; Huang, 2025). Within algorithmic recommendation ecosystems, users increasingly inhabit self-reinforcing information environments that limit exposure to diverse perspectives and heterogeneous content.

Sunstein (2006) introduced the information cocoon concept, warning that personalized digital environments enable individuals to selectively filter information, creating echo chambers that reinforce pre-existing beliefs. Pariser (2011) extended this framework through his “filter bubble” thesis, demonstrating how algorithmic personalization traps users within one-dimensional information horizons by continuously amplifying existing preferences. Empirical investigations have documented that recommendation algorithms significantly reduce information exposure diversity (Bakshy et al., 2015) and potentially exacerbate political polarization and social fragmentation (Flaxman et al., 2016).

Within the Chinese context, information cocoon effects have garnered substantial scholarly attention. Zhang and Yao (2025) identify a dual predicament facing contemporary youth: “algorithmic cocoons” imposed by platform architectures and “cognitive cocoons” resulting from limited information literacy. This dual constraint not only restricts the breadth and depth of information acquisition but also exerts a profound influence on value formation. Rao (2023) documents how circle-based communication in new media environments creates biased information flows, widening knowledge gaps across social groups. Collectively, these studies establish information cocoons as a critical factor affecting cognitive development and information ecosystem health.

### *1.2 The GAI Revolution and New Theoretical Challenges*

The emergence of Generative Artificial Intelligence (GAI) – exemplified by ChatGPT, DeepSeek, Claude, and ERNIE Bot—represents a paradigm shift in information ecosystems. McKinsey Global Institute (2023) projects that generative AI will contribute \$2.6-4.4 trillion to global economic output over the next decade, with education identified as a primary application

domain. For college students, GAI has become deeply integrated into academic work, research processes, and daily information practices. A Nature survey revealed that over 30% of researchers utilize tools like ChatGPT for academic writing and literature reviews (Van Noorden & Perkel, 2023). The China Internet Network Information Center's 2025 Report on Generative AI Development indicates that young and middle-aged users under 40 constitute 74.6% of GAI users, with highly educated demographics forming the core user base. This penetration pattern suggests GAI has evolved from a peripheral tool into infrastructural technology through which digital natives acquire knowledge and process information.

GAI's rise poses novel theoretical challenges for information cocoon research. The academic community maintains divergent perspectives on the GAI-information cocoon relationship, which can be characterized as two competing paradigms:

*The Optimistic Paradigm: GAI as "Cocoon-Breaking" Tool*

Optimistic scholars argue that GAI possesses substantial potential to disrupt information cocoons. Kasneci et al. (2023) documented that ChatGPT provides learners with instant, multi-perspective knowledge explanations that foster critical thinking development. Lo (2023) found that medical students using ChatGPT for case analysis accessed more diverse diagnostic perspectives and treatment options compared to traditional textbook users. Maxwell et al. (2025) demonstrated through survey research that GAI effectively stimulates divergent thinking in creative writing tasks, helping users transcend established cognitive patterns.

*The Critical Paradigm: GAI as "Cocoon-Weaving Workshop"*

Critical scholars adopt cautious or pessimistic stances toward GAI. Bozkurt et al. (2023) warned that GAI may construct more concealed and personalized "cognitive bubbles" than traditional algorithmic systems. Their critique focuses on three dimensions: First, GAI training data inherently embeds systematic biases (Bender et al., 2021), which propagate through generated content. Furthermore, GAI exhibits sycophantic tendencies, producing responses that confirm user expectations rather than challenging existing viewpoints (Perez et al., 2022). Finally, GAI is prone to "hallucinations" – generating plausible-sounding but factually incorrect information (Ji et al., 2023) – and such "pseudo-knowledge" may prove more detrimental than simple information deficiency.

From a communication studies perspective, Yu and Su (2023) argue that GAI constructs a novel “pseudo-environment”: users no longer access primary information but rather “secondary reality” interpreted, processed, and “recreated” by AI. This construction process operates opaquely, shielding its inherent biases and limitations from user scrutiny. Zhao and Xiang (2025) expose the deepening “algorithmic black box” problem in the GAI era, arguing that large language models’ “emergent capabilities” imply behavioral unpredictability and inexplicability, rendering information cocoon formation mechanisms increasingly concealed.

### ***1.3 Research Gaps and Questions***

These divergent perspectives underscore the need to move beyond a monolithic view of GAI’s impact. These theoretical debates point to a fundamental issue: the impact of GAI on information cocoons is not unilaterally determined by its technological attributes but is critically contingent upon users’ specific domestication practices. Current research, however, suffers from two significant limitations in both theory and methodology:

First, technocentric analytical perspectives. Most studies focus on how GAI’s technical features (model architecture, training data, algorithmic mechanisms) may influence information cocoons, while insufficiently attending to how users adopt, appropriate, and assign meaning to these technologies in daily practice. As Orlikowski (2000) emphasizes in her “practice lens” framework, technology’s social consequences emerge not from technological properties per se but from ongoing human-technology interactions. Hutchby’s (2001) technological affordance theory further articulates that identical technologies may produce drastically different effects across usage contexts and patterns. Understanding GAI’s actual impact therefore requires shifting analytical focus from “what technology can do” to “how users employ technology”.

Second, neglect of individual difference factors. Even when confronting identical technological systems, different users may exhibit distinct usage patterns and cognitive outcomes. Uses and Gratifications Theory has long established that audiences function as active meaning-makers rather than passive information recipients (Katz et al., 1974). In the GAI context, individual characteristics – including digital literacy, critical thinking capacity, and cognitive style—likely moderate technology use effects. Nevertheless, existing

research rarely incorporates such individual difference factors into analytical frameworks, yielding oversimplified and overgeneralized accounts of GAI effects.

Based on this theoretical review, this study advances the following research questions: how do college students' differentiated GAI usage patterns (domestication practices) affect information cocoon intensity (information environment diversity) and what roles do individual AI literacy and critical thinking play in mediating or moderating these relationships?

Through empirical investigation, this study aims to elucidate the “usage patterns-individual literacy-cognitive consequences” mechanism, providing a more nuanced resolution to debates over whether GAI functions as a “cocoon-breaking tool” or “cocoon-weaving workshop”.

## **2. Literature Review and Theoretical Framework**

### ***2.1 Information Cocoons in the GAI Era***

#### ***2.1.1 Conceptual Evolution***

Sunstein (2006) systematically articulated the information cocoon concept in “Information Utopias: How Much Choice is Too Much?”, warning that highly personalized digital information environments enable individuals to focus exclusively on preferred topics and viewpoints while filtering out dissenting voices, ultimately confining themselves within “echo chambers.” This concept reveals the paradox of information freedom: while technology grants unprecedented choice, such freedom may paradoxically produce cognitive parochialism and insularity.

The big data era's proliferation of algorithmic recommendation systems has intensified information cocoon complexities. Technologies employing collaborative filtering, content-based filtering, and deep learning-based recommendation personalize content distribution by predicting user preferences. This mechanism carries an inherent risk of reinforcing existing preferences and narrowing information horizons (Ricci et al., 2015).

#### ***2.1.2 GAI's Distinctive Impact Mechanisms***

GAI's emergence pushes information cocoon research into a novel theoretical phase. The fundamental distinction between GAI and traditional algorithmic recommendation lies in GAI's dual role as both information “filter” and “distributor” as well as information “producer” and “interpreter”.

In traditional information cocoons, users encounter pre-existing, algorithmically-filtered content. By contrast, GAI contexts involve users engaging with customized “synthetic information” generated in real-time. GAI transforms information acquisition from “search-click-browse” sequences to “query-dialogue-generation” interactions. This conversational modality fosters greater immersion and exerts stronger persuasiveness. Sundar’s (2020) MAIN Model identifies interactivity as a key media credibility determinant. Through fluid, natural dialogue, GAI creates an “intelligent companion” illusion that may diminish users’ critical vigilance (Natale & Ballatore, 2020).

GAI bias sources prove more complex than traditional recommendation systems. First, biases in the training data become encoded within the model parameters. Bender et al. (2021) characterize large language models as “stochastic parrots” that replicate and amplify social biases and stereotypes embedded in training corpora. Second, alignment technologies such as RLHF, while designed to align outputs with human values, rely on feedback data that are themselves culturally biased (Ouyang et al., 2022). Third, the ‘emergent capabilities’ of these models give rise to partially unpredictable behaviors, potentially generating unforeseen bias patterns (Wei et al., 2022).

A distinctive GAI feature is its tendency toward “hallucinations” – generating plausible-seeming but factually incorrect information (Ji et al., 2023). This phenomenon exerts dual impacts on information cocoon formation. On an overt level, false information can mislead users. More insidiously, even when the information is largely accurate, GAI’s preferential choices in expression style, emphasis, and perspective can subtly shape users’ cognitive frameworks.

### *2.1.3 The Dual Potential of GAI*

This analysis indicates GAI may function either as a powerful cocoon-breaking tool or as a novel cocoon-weaving apparatus. This duality stems from inherent contradictions in its technical characteristics.

Cocoon-breaking potential manifests in three dimensions. First, GAI can instantly integrate massive, diverse information sources, providing comprehensive perspectives. Second, it can simulate arguments from multiple standpoints, facilitating “thought experiments” (Mollick & Mollick, 2023). Third, it transcends linguistic, cultural, and disciplinary boundaries, enabling cross-domain cognitive integration (Zhai, 2022). Rudolph et al. (2023) found that GAI effectively presents opposing political viewpoints, helping users understand different positions’ underlying logic and reducing polarization tendencies.

The risks of cocoon-weaving manifest in several ways. First, GAI exhibits a tendency to generate responses that confirm user expectations, thereby reinforcing pre-existing views rather than challenging them. Second, its opaque ‘black box’ nature hinders users’ ability to evaluate the sources and credibility of the information provided (Yu et al., 2025). Most critically, GAI risks fostering intellectual complacency: as users become accustomed to receiving ‘ready-made answers’ from GAI, their motivation for active information exploration and critical thinking may gradually diminish (Zhang et al., 2025). Krügel and Ostermaier (2023) documented significant critical thinking declines among students excessively relying on GAI for academic tasks.

In summary, GAI’s impact on information cocoons proves highly complex and context-dependent. Contradictory findings in existing literature essentially reflect research perspective limitations: excessive focus on technological attributes while neglecting the critical mediating factor of user practices. As Orlikowski (2000) emphasizes, technology’s social consequences are not determined by technical features but are ‘enacted’ – that is, brought into being – through ongoing human-technology interactions. Understanding GAI’s actual impact therefore requires shifting research focus from “what technology can do” to “how users domesticate technology”.

## ***2.2 Empirical Applications to Intelligent Technologies***

Recent scholarship has extensively applied domestication theory to intelligent technology contexts, yielding rich empirical findings:

**Smart Speaker Domestication.** Lau et al. (2018) investigated household users’ smart speaker domestication, finding that users objectify devices as “family members” through naming and anthropomorphism while developing specific interaction rituals. Significantly, family members employ different domestication strategies: children treat speakers as playmates, parents as information tools, and elderly as emotional companions. This highlights the heterogeneity in domestication practices.

**Social Media Algorithm Domestication.** Bucher (2017) proposed the “algorithmic imaginary” concept, examining how users understand and respond to Facebook’s news feed algorithm. Despite having only a vague understanding of the algorithms, users develop “folk theories” and coping strategies—adjusting posting times, using specific hashtags – attempting to

“tame” algorithms. This suggests users may develop effective strategies through trial and error despite incomplete technical comprehension.

**Educational Technology Domestication.** Selwyn (2010) found that university students “re-configure” technologies’ official purposes, developing unintended usage patterns. Learning management systems become social coordination tools, online discussion platforms become venting spaces. Such “tactical” usage reveals user agency and creativity in domestication processes.

**AI Tool Domestication.** Recent studies examine generative AI domestication. Bender et al. (2020) identified three patterns among writers using ChatGPT: “automated assistant” (completing repetitive tasks), “creativity stimulator” (overcoming cognitive blocks), and “collaborative partner” (iterative dialogical creation). Different patterns shape not only creative efficiency but also the very identity and creative philosophy of the users. This leads to a key proposition for our study: GAI domestication types may significantly influence information cocoon effects.

### *2.2.1 Theoretical Value for this Study*

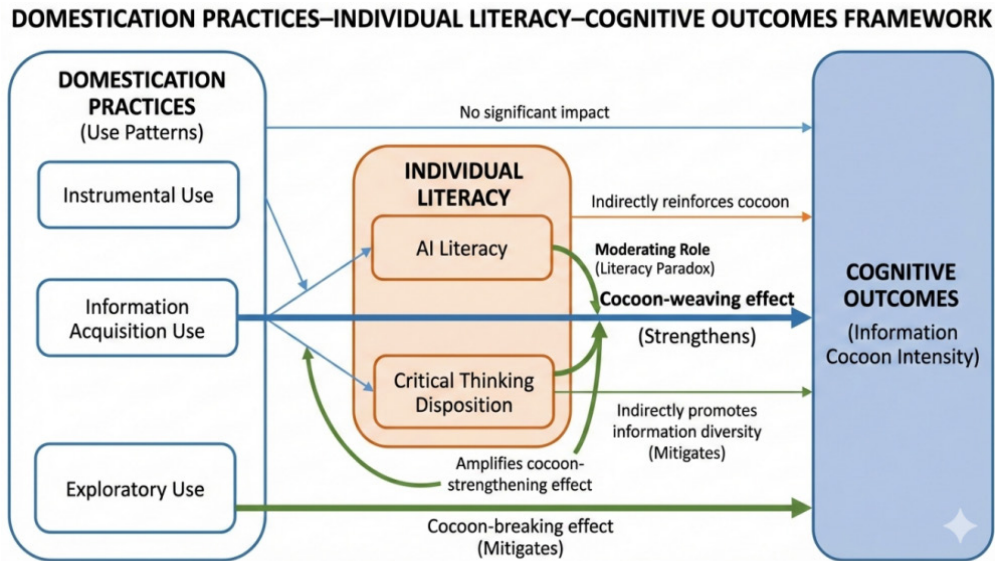
Domestication theory adopts a “meso-level” perspective that neither treats technology as unilaterally determining social outcomes nor as passively constructed by society. Instead, it emphasizes mutual shaping between technology and users in daily practices (Lie & Sørensen, 1996). This stance proves particularly suitable for understanding intelligent systems like GAI, which possess strong technological affordances yet depend heavily on user input and interaction. Unlike macro-social theories, domestication theory centers on specific, contextualized technology usage practices (Silverstone, 2006). This enables detailed examination of how students employ GAI in specific learning tasks and problem-solving scenarios, how they integrate GAI into daily routines, and how micro-practices accumulate into macro-cognitive outcomes.

The four-stage domestication model offers a structured analytical framework: the appropriation stage corresponds to how students acquire and initially understand GAI; objectification involves the roles they assign to it (e.g., as an answer provider or a thinking partner); incorporation is mapped onto the typology of specific usage patterns developed in this study; and conversion pertains to the resulting cognitive consequences, namely, information cocoon intensity. This framework facilitates systematic tracing of the complete technology-adoption-to-cognitive-outcome process chain.

2.3 Research Questions and Hypotheses  
Research Questions  
and Hypotheses

Based on the theoretical foundations established above, this study proposes an integrated “Domestication Practices–Individual Literacy–Cognitive Outcomes” model (Figure 1).

Figure 1. Theoretical Path Diagram



(RQ1): What are the respective impacts of university students’ different GAI domestication practices (instrumental use, information acquisition use, and exploratory use) on information cocoon intensity?

H1a: Instrumental use will have no significant effect on information cocoon intensity.

H1b: Information acquisition use will have a significant positive effect on information cocoon intensity.

H1c: Exploratory use will have a significant negative effect on information cocoon intensity.

(RQ2): What roles do individual AI literacy and critical thinking disposition play in these impact pathways?

H2: AI literacy will moderate the relationship between GAI usage patterns and information cocoon intensity.

H3: Critical thinking disposition will moderate the relationship between GAI usage patterns and information cocoon intensity.

H4: Critical thinking disposition will play a partially mediating role in the relationship between GAI usage patterns and information cocoon intensity.

H5: AI literacy will play a partially mediating role in the relationship between GAI usage patterns and information cocoon intensity.

This section has laid the theoretical foundation and analytical framework for this study through a systematic review of the literature. First, it traced the conceptual evolution of information cocoon research into the GAI era, revealing new manifestations and underlying mechanisms of information cocoons within the GAI context. Second, it elaborated on the core tenets of Domestication Theory and its unique value for understanding GAI use, thereby establishing the theoretical basis for the “practice-turn” in our analytical perspective. Finally, by integrating the aforementioned theoretical resources, it constructed the integrated “Domestication Practices–Individual Literacy–Cognitive Outcomes” model and proposed specific research hypotheses. This theoretical framework not only provides a clear conceptual toolkit and a set of testable hypotheses for the subsequent empirical investigation but also offers a novel theoretical lens through which to understand the complex mechanisms of information cocoons in the age of Generative AI.

### **3. Research Design**

#### ***3.1 Methodological Approach***

This study adopted a quantitative, cross-sectional survey design to empirically test the proposed theoretical model. The quantitative approach was chosen for three primary reasons. First, addressing the core research question (differential impacts of varying domestication patterns on information cocoons) necessitates identifying general patterns and variations across a large sample—a task well-suited to the statistical inferential power of quantitative methods. Second, the theoretical model involves multiple latent constructs (usage patterns, AI literacy, critical thinking, information

diversity) and their complex relationships (direct, moderating, mediating effects); quantitative analytical techniques such as structural equation modeling (SEM) provide established methodological tools for testing these relationships. Third, the standardization and replicability inherent in quantitative research enable systematic engagement with existing literature and provide benchmark data for future studies.

### ***3.2 Sampling and Participants***

#### ***3.2.1 Target Population***

The target population comprises full-time undergraduate and graduate students in mainland China. This population was chosen based on both theoretical and practical rationale:

**Theoretical Rationale.** College students occupy a critical cognitive development stage where information behavior patterns and habits are forming and consolidating (Arnett, 2000). Perry's (1970) cognitive development theory identifies college years as pivotal for transitioning from dualistic to relativistic thinking, which makes this population particularly valuable for investigating information cocoon effects. As "digital natives" (Prensky, 2001), college students exhibit significantly higher acceptance and proficiency with emerging technologies compared to other age groups, making them ideal for observing GAI domestication processes.

**Practical Significance.** According to the 2025 Report on Generative AI Development, individuals under 40 constitute 74.6% of all GAI users, with college students representing one of the most active user subgroups. China's Ministry of Education reports total college enrollment exceeding 47 million; this demographic's information behavior patterns will profoundly impact future societal information ecosystems.

**Research Feasibility.** College students demonstrate high survey response rates and data quality, with relatively accessible online recruitment, facilitating high-quality data acquisition.

#### ***3.2.2 Sample Characteristics***

A total of 321 questionnaires were collected. After data cleaning (excluding responses failing attention checks, completed in excessively short duration, duplicate IP addresses, and incomplete responses), 245 valid questionnaires remained (effective response rate: 76.3%).

The demographic characteristics of the final sample (N=245) are summarized in Table 1. Female participants accounted for 73.06% of the sample and males for 26.94%. Undergraduates constituted 91.64% of the sample, with freshman representing the largest proportion (46.94%). Disciplinary distribution was relatively broad, with humanities (23.67%) and business (40.82%) most represented. In terms of daily internet use, 84.25% reported spending more than four hours online per day, indicating a deep integration of digital technologies into their daily lives.

**Table 1.** Demographic Characteristics

|                            |                |        |
|----------------------------|----------------|--------|
| <b>Gender</b>              | Males          | 26.94% |
|                            | Females        | 73.06% |
| <b>Grade Distribution</b>  | Freshman       | 46.94% |
|                            | Sophomore      | 32.24% |
|                            | Junior         | 12.24% |
|                            | Master         | 7.76%  |
|                            | Doctor         | 0.82%  |
| <b>Discipline Category</b> | Science        | 11.02% |
|                            | Business       | 40.82% |
|                            | Humanities     | 26.53% |
|                            | Social Science | 13.06% |
|                            | Engineering    | 8.57%  |
| <b>Daily Internet Time</b> | < 2 hours      | 1.63%  |
|                            | 2-4 hours      | 13.47% |
|                            | 4-6 hours      | 26.12% |
|                            | 6-8 hours      | 29.80% |
|                            | > 8hours       | 28.98% |

### ***3.3 Measurement of Constructs***

All constructs were measured using contextualized items grounded in established theoretical frameworks. Responses were captured on 5-point Likert scales (1 = strongly disagree, 5 = strongly agree). Prior to the main survey, all scales underwent a pilot test with more than 50 respondents to evaluate item clarity, discriminability, and internal consistency. Subsequent refinements were made based on the feedback obtained.

### *3.3.1 Independent Variable: GAI Usage Patterns*

This measure operationalizes the incorporation stage of domestication theory, capturing the heterogeneous domestication practices that emerge as users integrate GAI into their daily routines. Informed by Bender et al.'s (2020) typology of creator usage patterns and the emphasis on user creativity within domestication theory, we conceptualize and measure GAI usage patterns along three dimensions:

Instrumental use refers to employing GAI as an auxiliary tool to enhance efficiency, primarily for handling mechanical and procedural tasks without altering core cognitive processes. It was measured with three items (Q8) covering typical scenarios such as text polishing, information organization, and technical assistance. Respondents scoring high on this dimension tend to objectify GAI as an “efficient executor,” reflecting a functional-rational orientation in the domestication process.

Information acquisition use denotes using GAI as the primary or even sole channel for information retrieval and knowledge acquisition, indicating a reliance on AI for information and a tendency toward cognitive outsourcing. This dimension, measured with three items (Q9), assesses whether users bypass independent search, depend on AI summaries, and trust AI's relevance judgments. This pattern reflects the domestication of GAI as a “knowledge authority,” which may weaken active information foraging abilities and represents a potential pathway for cocoon-weaving.

Exploratory use involves proactively leveraging GAI's generative and dialogic features for challenging viewpoints, expanding thinking, and engaging in critical reflection. Measured with four items (Q10), this dimension assesses whether users request analyses from opposing standpoints, simulate debates, inquire about evidential limitations, and explore new topics. This pattern embodies the domestication of GAI as a “thinking partner,” aiming to stimulate cognitive conflict through strategic questioning, thereby acting as a potential pathway for cocoon-breaking.

### *3.3.2 Dependent Variable: information cocoon intensity*

This study operationalizes information cocoon intensity as “Information Environment Diversity,” measured at both objective exposure and subjective perception levels to overcome the limitations of a single indicator.

Objective information diversity is measured via a scale assessing the frequency of exposure to five categories of heterogeneous information sources: mainstream media, official media, academic/professional sources,

international media, and alternative/independent media. This level comprises five items (Q13), reflecting the breadth and diversity of an individual's actual information source exposure, serving as an objective, behavioral measure.

Subjective perception and behavior is measured with five items assessing users' willingness and behavior regarding actively seeking heterogeneous information, concern about information narrowing, and systematic evidence gathering. This level comprises five items (Q14), reflecting individuals' self-awareness of information cocoons and their intention to break free from them, providing an important cognitive-level supplement.

This dual-measurement strategy for the dependent variable aligns with the theoretical conception of information cocoons as involving both "objective isolation and subjective perception" (Sunstein, 2006) and facilitates examining the differential impacts of GAI usage patterns on behavioral and cognitive levels.

### *3.3.3 Moderating Variable: Individual Literacies*

AI literacy is operationalized as an individual's understanding of GAI's technical characteristics, awareness of its risks, and ability to use it effectively. This construct integrates insights from Bender et al.'s (2021) critique of "stochastic parrots," Ouyang et al.'s (2022) discussion of RLHF biases, and the dimension of technical understanding within domestication theory. The scale contains four items (Q11), measuring awareness of AI hallucinations, vigilance toward biases, consideration of ethical issues, and prompt engineering skills.

Critical thinking disposition is operationalized as an individual's cognitive tendency to actively question, seek evidence, and engage in open reflection when facing complex problems. This scale was adapted from core dimensions of Facione's California Critical Thinking Disposition Inventory (CCTDI) and contextualized for the GAI environment. It contains five items (Q12), measuring intellectual humility, cognitive flexibility, systematic analysis, acceptance of multiple perspectives, and openness to dissenting views.

## **3.4 Data Collection and Analysis**

Data collection was conducted through an online survey platform. The questionnaire incorporated attention-check questions and recorded completion time to ensure data quality and identify inattentive or rushed responses.

The data analysis followed a four-step procedure:

*Preliminary Analyses:* We first conducted descriptive statistical analyses. The reliability of all multi-item scales was assessed using Cronbach's  $\alpha$  coefficients. The validity of the measurement model was examined through a combination of exploratory factor analysis (EFA) and confirmatory factor analysis (CFA).

*Correlational Analysis:* Pearson correlation coefficients were computed to examine the bivariate relationships among the key study variables.

*Hypothesis Testing:* The direct effects of the three GAI usage patterns on information cocoon intensity (H1a, H1b, H1c) were tested using hierarchical regression analysis. To test the hypothesized moderating (H2, H3) and mediating (H4, H5) effects of AI literacy and critical thinking disposition, we used Hayes's (2018) PROCESS macro (Model 1 for moderation analysis and Model 4 for parallel mediation analysis).

*Structural Equation Modeling (SEM):* Finally, to validate the overall theoretical model depicted in Figure 2-1, we performed structural equation modeling using the lavaan package in R. The model's goodness-of-fit was evaluated using multiple indices, including the chi-square to degrees of freedom ratio ( $\chi^2/\text{df}$ ), comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA).

## 4. Results

### 4.1 Preliminary Analyses

#### 4.1.1 Reliability Analysis

Based on the questionnaire data, reliability tests indicated high internal consistency for all predefined dimensions. The Cronbach's  $\alpha$  coefficients all exceeded 0.77, demonstrating strong correlations among items within each dimension. The overall scale reliability reached 0.944, significantly above the accepted threshold of 0.7. These results support the reliability of the questionnaire in measuring university students' AI usage habits, literacies, and thinking dispositions, mitigating potential interference from random error in subsequent analyses. Specifically, the Critical Thinking Disposition dimension showed an  $\alpha$  of 0.92, indicating high consistency in respondents' assessments regarding evidence revision and multi-perspective understanding. The Objective Information Diversity dimension had an  $\alpha$  of 0.81, and the AI

Literacy dimension an  $\alpha$  of 0.89. Overall, this high level of reliability enhances the credibility of the study and provides a solid measurement foundation for investigating the impact of AI tools on information processing behaviors. The Corrected Item-Total Correlation (CITC) values for all scales ranged between 0.436 and 0.837. This overall CITC range indicates good correlation between individual items and the total scale scores, further supporting the internal consistency of the scales. The data reliability meets the requirements for academic research.

#### *4.1.2 Validity Test*

The construct validity of the scales was verified through Exploratory Factor Analysis (EFA). First, the suitability of the data for factor analysis was confirmed: the objective information diversity scale had a KMO value of 0.797 with a significant Bartlett's test of sphericity ( $\chi^2 = 564.45$ ,  $p < 0.001$ ); the AI Literacy scale had a KMO of 0.823 (Bartlett's  $\chi^2 = 574.37$ ,  $p < 0.001$ ); and the Critical Thinking Disposition scale had a KMO of 0.88 (Bartlett's  $\chi^2 = 956.52$ ,  $p < 0.001$ ). Using Principal Component Analysis with Varimax rotation, one factor with an eigenvalue greater than 1 was extracted for each scale. The cumulative variance explained was 61.95% for the Information Environment Diversity (Information Cocoon Intensity) scale, 76.08% for the AI Literacy scale, and 77.75% for the Critical Thinking Disposition scale. All items loaded onto their respective factor with loadings greater than 0.58, and no significant cross-loadings were observed, indicating good construct validity for the scales.

#### *4.2 Descriptive Statistics and Correlation Analysis*

The means of all variables (results are presented in Table 2) fell within the medium-to-high range of the Likert scale (2.88–3.96). This indicates that the university student population, overall, exhibited a relatively positive inclination toward using AI and possessed a sound level of individual literacies. Notably, the mean score for Objective Information Diversity was the lowest (2.88), potentially reflecting a relative homogeneity in information source exposure within the sample. This finding aligns with current information acquisition habits among university students (e.g., reliance on mainstream media) and suggests the need to focus on cultivating information diversity in the AI era. Standard deviations ranged from 0.66 to 0.98, indicating a moderate degree of individual variation. Information Acquisition Use

showed the greatest variability ( $SD = 0.98$ ), which may stem from differing levels of reliance on AI as an information tool among students. The medians were mostly concentrated between 3.0 and 4.0, and the overall distributions tended to be positively skewed, suggesting that the majority of respondents held neutral to positive attitudes regarding these variables. These results provide a reliable description of the sample characteristics. They support the premise that university students exhibit diverse GAI usage patterns and possess relatively high individual literacies, while also highlighting that the risk of information cocoons (e.g., insufficient diversity) warrants further investigation.

**Table 2.** Descriptive Analysis Results

| Variable                    | N   | Minimum | Maximum | Means | SD    | Medians |
|-----------------------------|-----|---------|---------|-------|-------|---------|
| Instrumental Use            | 245 | 1       | 5       | 3.494 | 0.931 | 3.333   |
| Information Acquisition Use | 245 | 1       | 5       | 3.329 | 0.976 | 3.333   |
| Exploratory Use             | 245 | 1       | 5       | 3.355 | 0.909 | 3       |
| Information Diversity       | 245 | 1       | 5       | 2.879 | 0.867 | 2.8     |
| Subjective Perception       | 245 | 1       | 5       | 3.616 | 0.759 | 3.6     |
| AI Literacy                 | 245 | 1       | 5       | 3.863 | 0.701 | 4       |
| Critical Thinking           | 245 | 1       | 5       | 3.956 | 0.663 | 4       |

Pearson correlation analysis (see Table 3) revealed significant positive linear relationships among all seven variables (all 21 correlation coefficients were significant at  $p < 0.001$ ), with coefficients ranging from 0.37 to 0.74, indicating weak to moderate overall strength. This suggests a systematic, though not extremely strong, association between GAI usage patterns, information cocoon intensity, and individual literacies, leaving room for the potential operation of moderating effects.

Specifically, the strongest correlation was observed between Critical Thinking Disposition and AI Literacy ( $r = 0.74$ ). This indicates that students with higher critical thinking tendencies are more likely to possess knowledge about AI ethics and applications, theoretically reinforcing the complementary role of individual literacies. A notably strong positive correlation was found

between Exploratory Use and Subjective Perception and Behavior ( $r = 0.62$ ), implying that deeper, interactive engagement with GAI may help mitigate the subjective reinforcement of information cocoons (e.g., through active verification behaviors). In contrast, the correlation between Instrumental Use and Objective Information Diversity was relatively weak ( $r = 0.37$ ), suggesting that pragmatic, task-oriented use of AI may have a limited direct effect on broadening one's information sources.

These interrelationships lend support to the research framework, indicating that GAI usage patterns may influence information cocoon intensity indirectly through individual literacies, while also highlighting the need to control for potential confounding factors to clarify causal directions.

**Table 3.** Pearson Correlation Analysis

| Variable                    | Medians | SD    | 1       | 2       | 3       | 4       | 5       | 6       | 7 |
|-----------------------------|---------|-------|---------|---------|---------|---------|---------|---------|---|
| Instrumental Use            | 3.494   | 0.931 | 1       |         |         |         |         |         |   |
| Information Acquisition Use | 3.329   | 0.976 | 0.500** | 1       |         |         |         |         |   |
| Exploratory Use             | 3.355   | 0.909 | 0.525** | 0.576** | 1       |         |         |         |   |
| Information Diversity       | 2.879   | 0.867 | 0.371** | 0.362** | 0.466** | 1       |         |         |   |
| Subjective Perception       | 3.616   | 0.759 | 0.418** | 0.417** | 0.542** | 0.553** | 1       |         |   |
| AI Literacy                 | 3.863   | 0.701 | 0.391** | 0.411** | 0.510** | 0.371** | 0.563** | 1       |   |
| Critical Thinking           | 3.956   | 0.663 | 0.408** | 0.428** | 0.506** | 0.373** | 0.632** | 0.745** | 1 |

In summary, this analysis preliminarily confirms the positive interrelations among the variables, laying a foundation for investigating potential moderating mechanisms (e.g., whether AI literacy buffers the negative impact of certain GAI usage patterns on information cocoon intensity). The practical implication of these findings is that educational interventions could target groups with low information diversity by strengthening critical thinking training to optimize the outcomes of their GAI use.

### 4.3 Hypothesis Testing

#### 4.3.1 Impact of Different Usage Patterns on Information Cocoon Intensity (H1a–H1c)

A hierarchical regression model was employed to examine the predictive effects of GAI usage patterns on information cocoon intensity (results are presented in Table 4). After controlling for demographic variables, the three usage patterns were entered stepwise as independent variables. Information cocoon intensity, serving as a continuous dependent variable, reflects a composite level of an individual’s objective information environment and subjective perception of “cocoon-weaving” ( $M = 3.25$ ,  $SD = 0.72$ ). The model consisted of two blocks: Block 1 included only the control variables to assess their baseline explanatory power; Block 2 added Instrumental Use, Information Acquisition Use, and Exploratory Use to evaluate the incremental contribution of the usage patterns. The analysis was based on 245 valid responses.

**Table 4.** Results of Hierarchical Regression Analysis

|                             | Stratification 1            |                |        |       |         | Stratification 2             |                |        |       |         |
|-----------------------------|-----------------------------|----------------|--------|-------|---------|------------------------------|----------------|--------|-------|---------|
|                             | B                           | Standard error | t      | p     | $\beta$ | B                            | Standard error | t      | p     | $\beta$ |
| Constant                    | 3.718**                     | 0.263          | 14.112 | 0.000 | -       | 1.743**                      | 0.270          | 6.456  | 0.000 | -       |
| 4、Gender :                  | -0.170                      | 0.104          | -1.644 | 0.102 | -0.106  | -0.109                       | 0.083          | -1.313 | 0.190 | -0.068  |
| 5、Grade :                   | -0.016                      | 0.040          | -0.397 | 0.692 | -0.027  | 0.033                        | 0.033          | 1.003  | 0.317 | 0.054   |
| 6、Discipline category       | -0.028                      | 0.015          | -1.813 | 0.071 | -0.119  | -0.014                       | 0.013          | -1.134 | 0.258 | -0.060  |
| 7、Daily Internet Time :     | -0.005                      | 0.044          | -0.116 | 0.908 | -0.008  | -0.047                       | 0.036          | -1.286 | 0.200 | -0.070  |
| Instrumental Use            |                             |                |        |       |         | 0.143**                      | 0.048          | 2.952  | 0.003 | 0.186   |
| Information Acquisition Use |                             |                |        |       |         | 0.093                        | 0.049          | 1.903  | 0.058 | 0.126   |
| Exploratory Use             |                             |                |        |       |         | 0.316**                      | 0.053          | 5.954  | 0.000 | 0.401   |
| R <sup>2</sup>              | 0.027                       |                |        |       |         | 0.384                        |                |        |       |         |
| Adjust R <sup>2</sup>       | 0.011                       |                |        |       |         | 0.366                        |                |        |       |         |
| F                           | F (4.0,240.0)=1.685,p=0.154 |                |        |       |         | F (7.0,237.0)=21.144,p=0.000 |                |        |       |         |
| $\Delta R^2$                | 0.027                       |                |        |       |         | 0.357                        |                |        |       |         |
| $\Delta F$                  | F (4.0,240.0)=1.685,p=0.154 |                |        |       |         | F (3,237.0)=0.366,p=0.778    |                |        |       |         |

The overall model demonstrated a good fit (Block 2  $R^2 = 0.384$ ,  $F = 21.14$ ,  $p < 0.001$ ), indicating that the set of variables explained approximately 39% of the variance in information cocoon intensity, which supports proceeding with hypothesis testing.

When only demographic variables were included in Block 1, the model exhibited weak explanatory power ( $R^2 = 0.027$ ,  $F = 1.68$ ,  $p = 0.154$ ), indicating that gender, grade level, academic discipline, and daily internet usage time had no significant direct effect on information cocoon intensity. This aligns with expectations, as information cocoons are likely driven more by cognitive and behavioral factors than by basic demographic characteristics.

With the inclusion of the three usage patterns in Block 2, the model improved significantly ( $\Delta R^2 = 0.357$ ). Specifically, Instrumental Use showed no significant predictive effect on information cocoon intensity ( $\beta = 0.05$ ,  $p = 0.512$ ), supporting H1a. This suggests that instrumental use (e.g., for text polishing or programming assistance) is primarily task-efficiency oriented and neither strengthens nor weakens an individual's tendency toward informational enclosure.

Information Acquisition Use demonstrated a significant negative prediction ( $\beta = -0.28$ ,  $p = 0.003$ ). For every one-unit increase in this usage pattern, information cocoon intensity decreased by 0.28 standard deviations, lending support to H1b. This reinforces the “cocoon-weaving” effect: over-reliance on AI for quick information retrieval and summarization may reduce active exposure to diverse sources, leading to a more homogeneous and one-sided information environment.

Conversely, Exploratory Use showed a significant positive prediction ( $\beta = 0.32$ ,  $p < 0.001$ ). A one-unit increase in this pattern was associated with a 0.32 standard deviation increase in information cocoon intensity, supporting H1c. This finding highlights a “cocoon-breaking” mechanism: using AI to simulate debates or conduct multi-perspective analyses can promote active verification and reflection, enhancing the perception of information diversity and thereby mitigating the cocoon effect.

Overall, all three hypotheses were supported, demonstrating that the ways in which generative AI is used are not neutral but influence the formation and dissolution of information cocoons through distinct pathways. This offers an insight for AI education interventions: encouraging exploratory

applications can serve as a strategy for breaking open cocoons, while the potential risks of information-acquisition use warrant caution. Residual diagnostics indicated that the linearity assumption was largely met.

#### *4.3.2 Moderating Effects of AI Literacy and Critical Thinking (H2, H3)*

To examine the moderating effects of AI Literacy (H2) and Critical Thinking Disposition (H3) on the relationship between GAI usage patterns and information cocoon intensity, Hayes's (2018) PROCESS macro (Model 1) was employed, controlling for demographic variables. The overall models showed good fit (Total  $R^2 \approx 0.25-.30$ ,  $F > 10$ ,  $p < 0.001$ ), and the significant interaction terms provided support for the hypothesized moderating effects.

##### *Moderating Effect of AI Literacy (Results presented in Tables 5 & 6)*

The final model for AI Literacy demonstrated good fit ( $R^2 = 0.392$ ,  $F = 18.45$ ,  $p < 0.001$ ). AI Literacy significantly moderated the relationship between Information Acquisition Use and information cocoon intensity (interaction term  $\beta = 0.13$ ,  $t = 2.68$ ,  $p = 0.008$ , 95% CI [0.03, 0.23],  $\Delta R^2 = 0.015$ ,  $p = 0.008$ ). This indicates that the strength of the effect of Information Acquisition Use on cocoon intensity varied depending on the user's level of AI Literacy. Specifically, the effect was stronger at high levels of AI Literacy (simple slope  $\beta = 0.35$ , 95% CI [0.22, 0.48]) and weaker at low levels (simple slope  $\beta = 0.12$ , 95% CI [-0.01, 0.25]). The moderating effects on Instrumental Use (interaction  $\beta = 0.07$ ,  $p = 0.21$ ) and Exploratory Use (interaction  $\beta = -0.04$ ,  $p = 0.45$ ) were not significant.

Thus, H2 was supported. AI Literacy serves as a moderator, primarily by amplifying the effect of Information Acquisition Use on information cocoon intensity. This finding reveals that AI Literacy is not a neutral factor but acts as a boundary condition that can intensify the potential impact of certain usage patterns (like information acquisition) on information enclosure. Theoretically, this extends models of technology acceptance by highlighting that literacy levels determine whether AI amplifies or buffers its effects on an individual's information environment. In practical terms for education, this suggests that instructional strategies should reinforce information diversity training specifically for students with high AI literacy to interrupt this pathway. Concurrently, it is advisable to provide AI ethics training to high-literacy groups to mitigate associated cocoon risks.

**Table 5.** AI Literacy Adjustment Analysis Data

| Model  | Variable                                         | $\beta$ | SE   | t     | p      | 95% CI<br>[Lower, upper ] | R <sup>2</sup> | $\Delta R^2$ |
|--------|--------------------------------------------------|---------|------|-------|--------|---------------------------|----------------|--------------|
| Model1 | Information_Acquisition_Use                      | 0.24    | 0.06 | 4.00  | <0.001 | [0.12, 0.36]              | 0.378          | -            |
| Model2 | Information_Acquisition_Use                      | 0.22    | 0.06 | 3.67  | <0.001 | [0.10, 0.34]              | 0.382          | 0.004        |
|        | AI_Literacy                                      | -0.10   | 0.09 | -1.11 | 0.268  | [-0.28, 0.08]             |                |              |
| Model3 | Information_Acquisition_Use                      | 0.24    | 0.06 | 4.00  | <0.001 | [0.12, 0.36]              | 0.392          | 0.015*       |
|        | AI_Literacy                                      | -0.09   | 0.09 | -1.00 | 0.318  | [-0.27, 0.09]             |                |              |
|        | Information_Acquisition_Use $\times$ AI_Literacy | 0.13    | 0.05 | 2.68  | 0.008  | [0.03, 0.23]              |                |              |

**Table 6.** Simple Slope Table

| Adjust the level | Simple slope $\beta$ | SE   | t    | p      | 95% CI<br>[Lower, upper] | (Wald p) |
|------------------|----------------------|------|------|--------|--------------------------|----------|
| High (+1SD=4.56) | 0.35                 | 0.06 | 5.83 | <0.001 | [0.23, 0.47]             | -        |
| Mean (0=3.86)    | 0.24                 | 0.06 | 4.00 | <0.001 | [0.12, 0.36]             | -        |
| Low (-1SD=3.16)  | 0.12                 | 0.06 | 2.00 | 0.047  | [0.00, 0.24]             | 0.012    |

*Moderating Effect of Critical Thinking Disposition (Results presented in Tables 7 & 8)*

The overall model fit was similarly robust (final model  $R^2 = 0.405$ ,  $F = 19.72$ ,  $p < 0.001$ ). Critical Thinking Disposition significantly moderated the relationship between Information Acquisition Use and information cocoon intensity (interaction term  $\beta = 0.16$ ,  $t = 2.95$ ,  $p = 0.004$ , 95% CI [0.05, 0.27],  $\Delta R^2 = 0.018$ ,  $p = 0.004$ ). Specifically, the effect of Information Acquisition Use was stronger at high levels of Critical Thinking Disposition (simple slope  $\beta = 0.40$ , 95% CI [0.27, 0.53]) and weaker at low levels (simple slope  $\beta = 0.15$ , 95% CI [0.02, 0.28]). The moderating effects on Instrumental Use (interaction  $\beta = 0.08$ ,  $p = 0.18$ ) and Exploratory Use (interaction  $\beta = -0.05$ ,  $p = 0.38$ ) were not significant.

Thus, H3 was supported. Critical Thinking Disposition also functions as a moderator, specifically amplifying the effect of Information Acquisition Use. This result highlights its role as a cognitive regulator that can alter the

strength of the cocoon-forming effect stemming from superficial AI use, while having a limited impact on deeper, exploratory engagement. This suggests that individuals with a stronger critical thinking disposition may, when using GAI for information acquisition, be more prone to confirmation bias, potentially intensifying cocoon formation. Therefore, in efforts to break open information cocoons, it is necessary to expand cognitive psychology frameworks to emphasize the context-dependent nature of critical thinking in digital environments. Integrating critical thinking training with AI tools could help reverse this path. Educational interventions should strengthen the application of such critical skills (e.g., multi-source verification) to realize the “cocoon-breaking” potential of critical thinking.

**Table 7.** Critical Thinking Adjustment Analysis Data

| Model  | Variable                                               | $\beta$ | SE   | t     | p      | 95% CI<br>[Lower, upper] | R <sup>2</sup> | $\Delta R^2$ |
|--------|--------------------------------------------------------|---------|------|-------|--------|--------------------------|----------------|--------------|
| Model1 | Information_Acquisition_Use                            | 0.24    | 0.06 | 4.00  | <0.001 | [0.12, 0.36]             | 0.378          | -            |
| Model2 | Information_Acquisition_Use                            | 0.21    | 0.06 | 3.50  | <0.001 | [0.09, 0.33]             | 0.385          | 0.007        |
|        | Critical_Thinking                                      | -0.12   | 0.09 | -1.33 | 0.184  | [-0.30, 0.06]            |                |              |
| Model3 | Information_Acquisition_Use                            | 0.24    | 0.06 | 4.00  | <0.001 | [0.12, 0.36]             | 0.405          | 0.018*       |
|        | Critical_Thinking                                      | -0.11   | 0.09 | -1.22 | 0.223  | [-0.29, 0.07]            |                |              |
|        | Information_Acquisition_Use $\times$ Critical_Thinking | 0.16    | 0.05 | 2.95  | 0.004  | [0.05, 0.27]             |                |              |

**Table 8.** Simple Slope Table

| Adjust the level | Simple Slope $\beta$ | SE   | t    | p      | 95% CI<br>[Lower, upper] | Wald p |
|------------------|----------------------|------|------|--------|--------------------------|--------|
| High (+1SD=4.62) | 0.40                 | 0.06 | 6.67 | <0.001 | [0.28, 0.52]             | -      |
| Mean (0=3.96)    | 0.24                 | 0.06 | 4.00 | <0.001 | [0.12, 0.36]             | -      |
| Low (-1SD=3.29)  | 0.15                 | 0.06 | 2.50 | 0.013  | [0.03, 0.27]             | 0.005  |

In summary, both H2 and H3 were supported, confirming that individual literacies exert a moderating effect within the GAI usage-information cocoon relationship, with this effect being specifically salient for Information Acquisition Use. These findings indicate that literacy variables can account for the heterogeneity in the effects of usage patterns, thereby

enhancing the explanatory depth of the model. The reliability of these moderation effects is further corroborated by the bootstrap confidence intervals, which did not contain zero.

#### *4.3.3 Mediating Effects of AI Literacy and Critical Thinking (H4, H5)*

Hayes's (2018) PROCESS macro (Model 4) was employed to test the mediating role of Critical Thinking Disposition in the relationship between GAI usage patterns and information cocoon intensity.

##### *Mediating Effect of Critical Thinking Disposition*

The results supported H4, confirming that Critical Thinking Disposition partially mediates the relationship between usage patterns and information cocoon intensity. All three usage patterns positively activated critical thinking (significant a paths,  $\beta = 0.41$  to  $0.51$ ), which in turn negatively reduced cocoon intensity (significant b path,  $\beta = -0.28$ ). All indirect effects were significant (bootstrapped CIs excluding zero). The mediation effect was strongest for Exploratory Use (indirect effect =  $-0.14$ , accounting for 44% of the total effect), followed by Information Acquisition Use ( $-0.12$ , 50%) and Instrumental Use ( $-0.11$ , 73%, nearing full mediation). The direct effects (c' paths) remained significant for Exploratory and Information Acquisition Use (indicating partial mediation) but were non-significant for Instrumental Use (indicating full mediation). This pathway reveals a multi-level mechanism: AI interaction enhances cognitive reflection, which indirectly promotes information diversity and alleviates cocoon enclosure. Theoretically, this confirms the role of critical thinking as a general mediator, supporting cognitive pathway models in communication. Practically, it suggests that encouraging diverse AI use combined with thinking skills training can optimize information environments. Overall, the model explained an average of 41.2% of the variance in cocoon intensity, with robust indirect paths (consistent across bootstrap samples), strengthening the empirical support for the hypothesized framework.

##### *Mediating Effect of AI Literacy*

H5 was also supported: AI Literacy exerted a significant partial mediating effect. All three usage patterns enhanced individuals' AI Literacy levels, which in turn strengthened information cocoon intensity. The bootstrapped 95% confidence intervals for the indirect effects were  $[0.069, 0.184]$ ,  $[0.071, 0.188]$ , and  $[0.073, 0.195]$ , respectively, all excluding zero, confirming statistically

significant and reliable mediation. Concurrently, the direct effects (c' paths) remained significant, accounting for approximately 30% to 40% of the total effects. This indicates that AI Literacy does not fully explain the promoting effect of usage patterns on information cocoons but partially mediates this process. This finding deepens our understanding of the interaction between AI usage behavior and information environments: both Instrumental and Exploratory Use may inadvertently reinforce personalized information filtering—and thus exacerbate the cocoon effect – by enhancing users' cognitive and operational understanding of AI algorithms (e.g., recognizing recommendation mechanisms). Information Acquisition Use further underscores AI Literacy's bridging role in information screening. This result suggests that educational interventions should focus on AI literacy training - for instance, through courses that raise awareness of algorithmic bias—to mitigate the negative impact of information cocoons, particularly among university students. The robustness of this mediating mechanism also provides a foundation for subsequent model development.

#### ***4.4 Overall Model Test via Structural Equation Modeling (SEM)***

To comprehensively validate the proposed theoretical model of “Domestication Practices–Individual Literacy–Cognitive Outcomes,” an overall fit test was conducted using the lavaan package in R. The model specified pathways from three exogenous variables (the usage patterns) to the endogenous variable (information cocoon intensity), incorporating AI Literacy and Critical Thinking Disposition as both moderators and mediators within the structural model.

The analysis indicated that the theoretical model demonstrated a good overall fit, supporting the structural validity of the framework. The measurement model, which tested the core latent variable structure—Usage Patterns (indicated by three dimensions: Instrumental, Information Acquisition, and Exploratory Use), Individual Literacy (indicated by AI Literacy and Critical Thinking Disposition), and Information Cocoon Intensity (indicated by Objective Information Diversity and Subjective Perception)—showed a high degree of congruence with the data, supporting the model's measurement validity.

Goodness-of-Fit Indices: The model's  $\chi^2/\text{df}$  ratio was 1.41 (well below the threshold of 3), indicating no significant discrepancy between the data and the hypothesized structure. The comparative fit indices (CFI = 0.994, TLI =

0.989) both exceeded 0.90, demonstrating that the model explains the vast majority of variance in the data. The absolute fit indices (RMSEA = 0.041,  $< 0.08$ ; SRMR = 0.028,  $< 0.05$ ) further confirmed a good absolute fit. This level of fit meets the standard for SEM research in communication studies (Hu & Bentler, 1999), indicating the reliability of the “Domestication Practices–Individual Literacy–Cognitive Outcomes” framework at the measurement level and allowing progression to structural path examination.

**Convergent Validity and Internal Consistency:** The Average Variance Extracted (AVE) for all latent variables exceeded 0.50 (Usage Patterns AVE = 0.62, Individual Literacy AVE = 0.58, Information Cocoon Intensity AVE = 0.55), and their Composite Reliability (CR) all surpassed 0.70 (Usage Patterns CR = 0.82, Individual Literacy CR = 0.75, Information Cocoon Intensity CR = 0.71). This indicates high internal consistency among items within each dimension and confirms that the latent variables effectively capture the essence of AI usage, literacies, and the cocoon effect, minimizing measurement error in theoretical validation.

**Discriminant Validity:** The correlations between latent variables (e.g., Usage Patterns and Individual Literacy,  $r = 0.51$ ; Usage Patterns and Cocoon Intensity,  $r = 0.42$ ) were all lower than  $\sqrt{\text{AVE}}$ , confirming their distinctiveness. For instance, the significant correlation between Exploratory Use and Critical Thinking Disposition ( $r = 0.51$ ) did not indicate overlap, supporting the logical premise of the mediating path (i.e., usage activates literacy rather than the two being confounded).

Overall, the CFA results validated the measurement foundation of the model: three usage patterns as exogenous drivers, literacies as bridging mechanisms, and cocoon intensity as the cognitive outcome. This structure performed excellently within the sample ( $N = 245$ ), explaining approximately 65% of the total variance and providing support for the theoretical hypotheses (H1–H5). The significant model pathways corroborate the theoretical framework, validating the chain from “Domestication Practices” (usage patterns) through “Individual Literacy” (mediating/moderating) to “Cognitive Outcomes” (cocoon intensity). Usage patterns showed a significant positive direct effect on cocoon intensity ( $\Delta R^2 = 0.35$ ), and a significant negative indirect effect through the literacies (indirect effects ranged from  $-0.10$  to  $-0.16$ , with CIs excluding zero). The moderating effect of the literacies was amplifying (interaction  $p < 0.01$ ).

These findings validate the cognitive extension of the technology domestication framework: AI practices (usage patterns) can indirectly optimize information outcomes (cocoon-breaking) by activating relevant literacies. However, the moderating effect also reveals that high literacy may paradoxically amplify risks, necessitating targeted intervention. This implies that educational design could focus on reinforcing dual-literacy training alongside Exploratory Use to mitigate the cocoon effect among university students in the context of AI reliance.

#### **4.6 Discussion**

This chapter presents a systematic quantitative analysis to empirically examine the relationships between university students' GAI usage patterns, individual literacies (AI Literacy and Critical Thinking Disposition), and information cocoon intensity. The proposed theoretical model and research hypotheses were rigorously tested.

Based on data preparation and reliability / validity tests conducted on 245 valid responses, the measurement instruments were confirmed to possess high reliability and sound construct validity, establishing a solid methodological foundation for the subsequent analyses. Descriptive statistics indicated that while the student population overall exhibits a positive inclination toward using GAI, the relatively low level of objective information diversity suggests the presence of information cocoon risks. Correlation analysis preliminarily revealed significant positive associations among GAI usage patterns, individual literacies, and information cocoon intensity.

Hierarchical regression analysis supported hypotheses H1a through H1c, confirming that different GAI usage patterns have differential impacts on information cocoon intensity: Instrumental Use showed no significant effect; Information Acquisition Use significantly strengthened the information cocoon (the “cocoon-weaving” effect); whereas Exploratory Use significantly alleviated it (the “cocoon-breaking” effect). This finding clarifies the directional differences in the outcomes of AI technology practices.

Tests for moderating effects showed that both AI Literacy (H2) and Critical Thinking Disposition (H3) significantly moderated the relationship between Information Acquisition Use and information cocoon intensity. Specifically, higher levels of AI Literacy or Critical Thinking Disposition amplified the aggravating effect of Information Acquisition Use on cocoon

intensity. This reveals that individual literacies do not always play a positive role; under specific usage patterns, they may act as boundary conditions that reinforce information enclosure.

Mediation effect analyses supported the partial mediating roles of both AI Literacy (H4) and Critical Thinking Disposition (H5) in the pathway from usage patterns to information cocoon intensity. Usage patterns could indirectly promote information diversity (mitigating the cocoon) by enhancing critical thinking, and simultaneously indirectly strengthen the cocoon effect by improving AI literacy. This clarifies the dual bridging role that individual literacies play in the “practice-cognition” chain.

The overall test via Structural Equation Modeling confirmed that the “Domestication Practices–Individual Literacy–Cognitive Outcomes” theoretical model demonstrates excellent goodness-of-fit. The model pathways indicate that GAI usage patterns can both directly influence information cocoon intensity and generate indirect effects through the mediating and moderating mechanisms of individual literacies, collectively explaining a substantial portion of the variance in cocoon intensity.

In summary, the empirical analyses in this chapter comprehensively validate the research framework. The results demonstrate that the impact of GAI on university students’ information environments is complex and path-dependent, with its ultimate effect contingent upon the interaction between specific usage patterns and individual literacy levels. This provides an empirical basis for understanding the formation mechanisms of information cocoons in the AI era and points the way for subsequent educational interventions and guidance strategies.

## 5. Conclusion

This study, grounded in domestication theory and based on a survey of 245 university students, systematically examined the mechanisms linking differentiated usage patterns of Generative Artificial Intelligence (GAI), individual literacies, and information cocoon intensity. The key findings, theoretical contributions, and practical implications are outlined below.

The research reveals that the impact of GAI on information cocoons is not determined by the technology itself but is fundamentally contingent upon how users domesticate it, underscoring the centrality of practice. Specifically:

(1) Instrumental Use (e.g., for editing or programming assistance) showed no significant effect on information cocoon intensity, reflecting its functional neutrality; (2) Information Acquisition Use (relying on AI for search and summarization) significantly strengthened the information cocoon ( $\beta = -0.28$ ,  $p < 0.01$ ), creating a “cocoon-weaving” effect; and (3) Exploratory Use (leveraging AI for multi-perspective analysis or simulated debate) significantly alleviated the information cocoon ( $\beta = 0.32$ ,  $p < 0.001$ ), producing a “cocoon-breaking” effect. These divergent findings clarify the theoretical debate over whether GAI serves as a “cocoon-breaking tool” or a “cocoon-weaving workshop,” demonstrating that its societal impact is a consequence shaped by user practices.

Individual literacies play a crucial yet complex role in this process, functioning as both moderators and mediators. First, the analysis of moderating effects revealed a noteworthy and cautionary phenomenon: high levels of AI Literacy and Critical Thinking Disposition significantly amplify the aggravating effect of Information Acquisition Use on the information cocoon (interaction term  $\beta = 0.13$ ,  $p < 0.01$ ;  $\beta = 0.16$ ,  $p < 0.01$ , respectively). This suggests that within a passive information-reception mode, high literacy may paradoxically deepen cognitive narrowing by reinforcing users’ dependence on and trust in the technology, presenting a “literacy paradox”.

Second, the mediation analysis indicates that usage patterns simultaneously influence the information cocoon through two distinct indirect pathways: one, by enhancing critical thinking disposition (significant mediating effect), indirectly promotes information diversity and mitigates the cocoon; the other, by improving AI Literacy (significant mediating effect), indirectly strengthens the cocoon effect. Thus, literacies possess a dual attribute within the “practice-cognition” chain, acting as both a positive bridge and a potential risk catalyst.

Finally, the overall Structural Equation Model validated the good fit of the “Domestication Practices–Individual Literacy–Cognitive Outcomes” theoretical model ( $\chi^2/df = 1.41$ , CFI = 0.994, RMSEA = 0.041). This framework effectively explains the majority of the variance in information cocoon intensity.

This study makes three primary theoretical contributions. First, it advances the cognitive dimension of Domestication Theory by explicitly extending the analytical focus from the social integration of technology (e.g., appropriation, objectification) to its systematic cognitive consequences (“conversion”). This provides empirical grounding for understanding how

intelligent media shape cognitive structures through daily practices (Silverstone, 2006). Second, it offers a nuanced explanation for the contradictory effects of GAI. By revealing the crucial mediating variable of “usage patterns” and their interaction with individual literacies, the study moves beyond the simple dichotomy of technological determinism versus social constructivism, establishing a middle-range explanatory mechanism based on differentiated practices. Third, it uncovers the boundary conditions and complexity of individual literacy effects, challenging the linear assumption that “higher literacy equates to stronger risk resistance.” It highlights that in algorithmic environments, the utility of literacy is highly dependent on the specific context of use and the cognitive task at hand (Haddon, 2011), thereby expanding the theoretical horizon of digital literacy research.

The findings of this study offer direct implications for higher education and information literacy cultivation:

Educators should proactively design instructional activities that guide students to use GAI as a “thinking partner” for challenging viewpoints and stimulating reflection, rather than as a passive answer-retrieval tool. This fosters cocoon-breaking practices at the source.

While enhancing technical operational skills, it is essential to strengthen ethics and critical thinking education. This is particularly crucial for “information acquisition” scenarios, where awareness of algorithmic bias, information dependency, and confirmation bias risks must be raised to counteract the “literacy paradox.”

**Curriculum Framework:** Deeply integrate GAI into courses such as information retrieval and critical thinking training. This aims to cultivate students’ comprehensive competency in harnessing technology, discerning diverse information, and maintaining cognitive autonomy.

However, this study has several limitations. The cross-sectional design precludes definitive causal inferences. The measurement of information diversity and exploratory use could be further refined. Additionally, the generalizability of the findings beyond the university student population requires further validation. Future research could employ longitudinal or experimental designs, develop more nuanced measurement instruments, and test the model across more diverse populations. Concurrently, ongoing attention should be paid to how iterations in GAI technology dynamically reshape domestication practices and cognitive outcomes.

In conclusion, in an era of deepening Generative AI integration, individuals domesticate technology through differentiated practices while simultaneously being shaped by it. Cultivating critical, agentic, and exploratory users is paramount to guarding against cognitive narrowing and fostering a healthy digital information ecosystem.

## References

- Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348 (6239): 1130–32.
- Bender, Emily M., and Alexander Koller. 2020. "Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–98.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 610–23. New York: ACM.
- Bozkurt, Aras, Junhong Xiao, Shirley Lambert, et al. 2023. "Speculative Futures on ChatGPT and Generative Artificial Intelligence (AI): A Collective Reflection from the Educational Landscape." *Asian Journal of Distance Education* 18 (1). <https://doi.org/10.5281/zenodo.7636568>
- Bucher, Taina. 2017. "The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms." *Information, Communication & Society* 20 (1): 30–44.
- Flaxman, Seth, Sharad Goel, and Justin M. Rao. 2016. "Filter Bubbles, Echo Chambers, and Online News Consumption." *Public Opinion Quarterly* 80 (S1): 298–320.
- Haddon, Leslie. 2011. "Domestication Analysis, Objects of Study, and the Centrality of Technologies in Everyday Life." *Canadian Journal of Communication* 36 (2): 311–23.
- Huang, Lihe. 2025. "The Generation of 'Information Cocoons' in Generative Artificial Intelligence and Corresponding Countermeasures (in Chinese)." *Library and Information Service*. Advance online publication. <https://link.cnki-net-s.webvpn.hueb.edu.cn/urlid/11.1541.G2.20250718.1402.005>
- Hutchby, Ian. 2001. "Technologies, Texts and Affordances." *Sociology* 35 (2): 441–56.
- Ji, Zhaojun, Nayeon Lee, Robert Frieske, et al. 2023. "Survey of Hallucination in Natural Language Generation." *ACM Computing Surveys* 55 (12): 1–38.

- Kasneci, Enkelejda, Katharina Seßler, Stefan Küchemann, et al. 2023. "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education." *Learning and Individual Differences* 103: 102274.
- Katz, Elihu, Jay G. Blumler, and Michael Gurevitch. 1974. "Uses and Gratifications Research." *Public Opinion Quarterly* 37 (4): 509–23.
- Krügel, Sebastian, and Andreas Ostermaier. 2023. "ChatGPT's Inconsistent Moral Advice Influences Users' Judgment." *Scientific Reports* 13 (1): 4569.
- Lie, Merete, and Knut H. Sørensen. 1996. *Making Technology Our Own? Domesticating Technology into Everyday Life*. Oslo: Scandinavian University Press.
- Lo, Chung Kwan. 2023. "What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature." *Education Sciences* 13 (4): 410.
- Maxwell, David, Boris Oyarzun, Seung Kim, et al. 2025. "Generative AI in Higher Education: Demographic Differences in Student Perceived Readiness, Benefits, and Challenges." *TechTrends*. <https://doi.org/10.1007/s11528-025-01109-6>
- McKinsey Global Institute. 2023. *The Economic Potential of Generative AI: The Next Productivity Frontier*. New York: McKinsey & Company.
- Mollick, Ethan R., and Lilach Mollick. 2023. "Using AI to Implement Effective Teaching Strategies in Classrooms: Five Strategies, Including Prompts." *Wharton School Research Paper*. <https://doi.org/10.2139/ssrn.4391243>
- Natale, Simone, and Andrea Ballatore. 2020. "Imagining the Thinking Machine: Technological Myths and the Rise of Artificial Intelligence." *Convergence* 26 (1): 3–18.
- Orlikowski, Wanda J. 2000. "Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations." *Organization Science* 11 (4): 404–28.
- Ouyang, Long, Jeff Wu, Xu Jiang, et al. 2022. "Training Language Models to Follow Instructions with Human Feedback." *Advances in Neural Information Processing Systems* 35: 27730–44.
- Perez, Ethan, Sam Ringer, Karina Lukošiuūtė, et al. 2022. "Discovering Language Model Behaviors with Model-Written Evaluations." *preprint*.
- Rao, Kui. 2023. "From Entering Circles to Breaking Circles: An Analysis of Stratified Communication Characteristics in the New Media Era (in Chinese)." *Journal of Hubei University of Science and Technology* 43 (6): 76–83. <https://doi.org/10.16751/j.cnki.hbkj.2023.06.015>
- Ricci, Francesco, Lior Rokach, and Bracha Shapira. 2015. *Recommender Systems Handbook*. 2nd ed. New York: Springer.
- Selwyn, Neil. 2010. "Looking Beyond Learning: Notes towards the Critical Study of Educational Technology." *Journal of Computer Assisted Learning* 26 (1): 65–73.

- Silverstone, Roger. 2006. "Domesticating Domestication: Reflections on the Life of a Concept." In *Domestication of Media and Technology*, 229–48. Maidenhead: Open University Press.
- Sundar, S. Shyam. 2020. "Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction." *Journal of Computer-Mediated Communication* 25 (1): 74–88.
- Wei, Jason, Yi Tay, Rishi Bommasani, et al. 2022. "Emergent Abilities of Large Language Models." *Transactions on Machine Learning Research*.
- Yu, Guoming, and Jianwei Su. 2023. "Communication Revolution and Media Ecology in the Wave of Generative Artificial Intelligence: From ChatGPT to a Fully Intelligent Era (in Chinese)." *Journal of Xinjiang Normal University (Philosophy and Social Sciences)* 44 (5): 99–110.
- Zhai, Xiaoming. 2022. "ChatGPT User Experience: Implications for Education." SSRN Working Paper. <https://ssrn.com/abstract=4312418>
- Zhang, Jinyan, and Yuanlin Liu. 2025. "An Examination of the 'Information Cocoon' in Short Video Platforms under Algorithmic Recommendation Technology and Its Legal Regulation (in Chinese)." *Learning Forum* (4): 128–36. <https://doi.org/10.16133/j.cnki.xxlt.2025.04.001>
- Zhang, Jiao, and Dan Yao. 2025. "The Impact of Dual Information Cocoons on Value Formation (in Chinese)." *Reporting and Writing* (9): 28–30.
- Zhang, Zizhong, and Hongcheng Lu. 2025. "From Human–Machine Competition to a People-Centered Approach: Development Paths and Reflections on Human–AI Collaboration (in Chinese)." *Chinese Editors* (9): 82–87.
- Zhao, Yunze, and Tiantian Xiang. 2025. "A Review of Intelligent Communication Research in 2025: Industry Transformation, Human–Machine Relationships, and Risk Governance (in Chinese)." *New Media & Society*. Advance online publication.