



HAL
open science

Datafying diplomacy: How to enable the computational analysis and support of international negotiations

Florian Cafiero

► **To cite this version:**

Florian Cafiero. Datafying diplomacy: How to enable the computational analysis and support of international negotiations. *Journal of computational science*, 2023, 71, pp.102056. 10.1016/j.jocs.2023.102056 . hal-04121762

HAL Id: hal-04121762

<https://hal.science/hal-04121762v1>

Submitted on 8 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Datafying diplomacy: How to enable the computational analysis and support of international negotiations

Florian Cafiero

médialab, Sciences Po, Paris, France

ARTICLE INFO

Keywords:

Computational diplomacy
Open data
Digital humanities

ABSTRACT

Computational diplomacy, the application of digital and computational methods to the study and practice of international negotiations, presents unique challenges compared to most other fields in the computational humanities and social sciences. Among them are the necessity of responsiveness when handling crises, the need to anticipate and respond to adversarial behavior, or the need for secrecy in dealing with sensitive information. In this paper, we propose an agenda to address these challenges, and evaluate the feasibility of the various tasks that could be assigned to computational diplomacy. While most analysis tools seem almost ready to use, the availability and reliability of diplomacy-related data remains a major concern.

1. The specificities of computational diplomacy

The recent emergence of computational diplomacy might appear to be a new but well-tracked approach. The application of digital then computational methods to the humanities and the social sciences have led to the rapid development of computational humanities [1] and computational social science [2] in the past few decades. Many challenges that computational diplomacy should face are thus likely to have been mostly addressed elsewhere, and could be tackled with the help of researchers from these fields. "Computational International Relations" [3], a scholarly field studying the history and political science of past and current international relations, is actually considered as belonging to the computational social sciences, and should obviously share objects and methods with Computational Diplomacy. Some of the tasks that Computational Diplomacy will have to perform have to do with Computational International Relations. Even if, as Vincent Pouliot and Jérémie Cornut stated, "diplomatic studies have long been the poor child of International Relations (IR) theory" [4], they still belong to this field, and are attracting an ever-growing attention. Going beyond interesting but punctual historical relations of a specific event, or testimonies and texts by practitioners has become the ambition of IR specialists in the past decade, who are trying to build larger theories about diplomacy [5]. The application of computational methods to diplomatic studies could help researchers to get a grasp on wider-scale or longer-term phenomena, thus accompanying this new movement in diplomatic studies. This would be no different that what happened in other fields in the humanities and social sciences such as literature [6], history [7] etc.,

which took advantage of new technical possibilities to develop or consolidate broader theories.

But it is obvious that we expect different things from Computational Diplomacy than a purely academic production. Diplomacy is not a scientific discipline per se, but rather a practice, a foreign policy tool among others, such as economic sanctions or outright war. And while the military [8,9] or economic [10] institutions already largely use computational methods to help their decisions, diplomacy seems to lie behind.

Methods developed by Computational Diplomacy and the results they yield should thus be usable to support the work of diplomats and decision-makers. Among the goals that can be set for this new discipline are tasks such as understanding and predicting the behavior of other states, the possible outcomes of a crisis, supporting diplomatic negotiations by helping to analyze complex situations, or monitoring the enforcement of international agreements.

Beyond the variety and huge amount of data needed, the challenge in collecting it for diplomatic decision-making is ensuring that the data is accurate, reliable, and up-to-date. This can be difficult because diplomatic issues often entail sensitive information and involve multiple stakeholders with different perspectives and agendas. Three major differences could thus differentiate computational diplomacy from a large part of the computational humanities and social sciences:

1. **The possibility of adversarial behavior:** not every document analyzed in computational diplomacy is made to deceive. But diplomacy obviously involves a large part of strategy. Assessing the

E-mail address: florian.cafiero@sciencespo.fr.

<https://doi.org/10.1016/j.jocs.2023.102056>

Received 9 December 2022; Received in revised form 6 March 2023; Accepted 29 April 2023

Available online 11 May 2023

1877-7503/© 2023 The Author. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

- authenticity of some documents, or the presence of false information is thus crucial to the reliability of the whole analysis.
2. **The need for speed:** handling an international crisis often requires quick responses, and even a slight delay in action can have dramatic consequences. The processing time cannot be the same as for a purely scholarly publication in computational social science or digital humanities. Having already the majority of the databases needed for our study ready should thus be a major preoccupation. Among those databases are numerical data, but also a huge amount of texts, written by the diplomats themselves, as well as by the governments involved in an international disputes, or by their citizens.
 3. **Secrecy:** Many documents that would be invaluable for analyses are simply not available to researchers. The articles 24 ("The archives and documents of the mission shall be inviolable at any time and wherever they may be. ") and 27 (regarding the diplomatic bag and diplomatic courier) of the 1961 Vienna Convention on Diplomatic Relations [11] protect on principle this secrecy. Beyond governments legitimate claims to secrecy for actual safety or strategic reasons, other parties involved in diplomatic negotiations can decide not to give access to some data or documents to researchers, fearing that some strategic information would be inferred about them, even from less sensitive documents.

In this regard, Computational Diplomacy has a lot in common with public health research, where medical secrecy limits on principle some

access to information, where adversarial behavior is to be expected by some parties involved, and for which timely outputs can be more than necessary. Making texts available, editing them properly and according to standards, authenticating them and identifying or profiling their author(s): those actually are the main tasks of the philologist. In that regard, a lot of what has to be done in the next years in the field of computational diplomacy should thus use or adapt a lot of the tools developed for "computational philology". Based on our experience with these two fields, are we already prepared to handle computational diplomacy? We imagine here a series of steps through which a diplomacy-related document should go through (Fig. 1) before being used for a computational diplomacy research, and try to assess the difficulties to come.

2. How ready are our sources ?

Many sources necessary for Computational Diplomacy can be collected through procedures usual elsewhere in Computational Social Science. If we want to analyze negotiations as "two-level games" [12], both international and international, we might be tempted to gather information on social movements or public opinion from social networks data, or socio-economic data from surveys and public databases. That kind of data could prove essential to the analysis of an international negotiation, to assess its expected profitability, or to understand the chance that the population and various actors involved beyond govern-

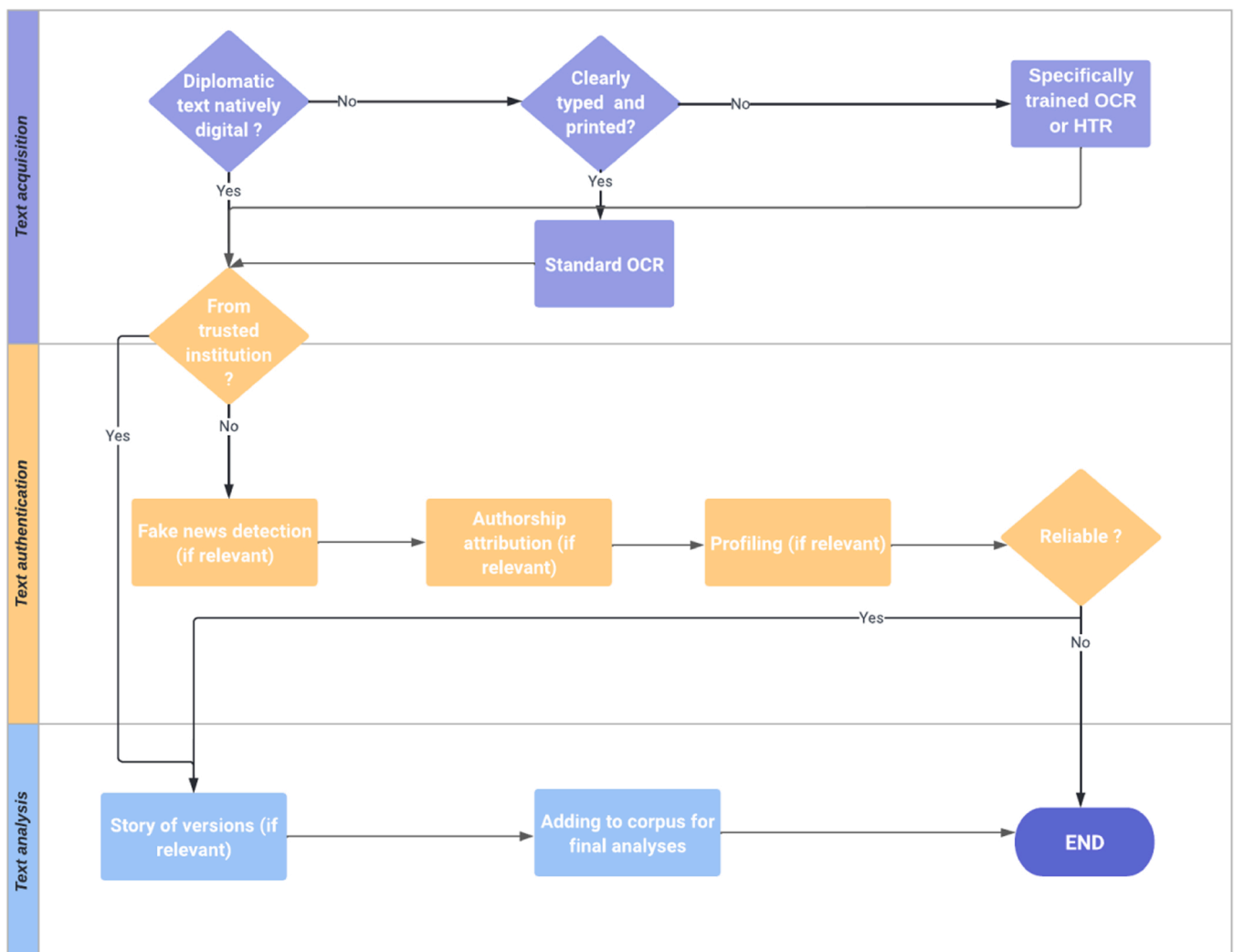


Fig. 1. Computational philology for diplomacy: a proposed workflow.

ments will or not accept its outcome. But the challenges of gathering and analyzing these data are well-known. We thus focus here on sources specific to the field.

2.1. Documents from States and International Organizations

Following a wider effort for archives' digitization, institutions linked to diplomacy have made available a more and more important number of documents online. The number and the nature of these documents differs a lot, but even the countries which invested the most do not provide sufficient sources for easily performed computational analyses.

In France, the *Ministère des Affaires des Etrangères* created with the French National Library's digital initiative *Gallica* a "Bibliothèque diplomatique numérique", opened in 2018. It encompasses a large collection coming from both the French National Library and the Foreign Affairs Ministry: treaties, conventions, national and intergovernmental official publications, but also a lot of documents regarding diplomatic history or diplomats themselves. In Canada for instance, the numerous sources available online on the "Global Affairs Canada Digital Library" are searchable PDFs, which allows for good information retrieval online. These more than 30.000 online pieces of archive are rarely available for the most recent years (one match only for documents coming from the past two years). The reason for this is probably two-fold: possible confidentiality issues, but more generally, the time to process documents which are not fed directly in a digital form, but rather printed then scanned.

Scanned documents shared as PDFs are also the norm for the USA, Swiss, UN digital archives. Unfortunately, that kind of documents adds a lot of pre-processing work that can be extremely time-consuming. They raise a large number of problems, from poor / old digitization, poorly opened book deforming the transcripts of parliamentary debates [13], stamps over the text, complex templates [14], manuscript notes in the margins etc. Getting around these apparently simple problems can become a nightmare for researchers in general, but all the more unbearable if we think of large-scale and fast operations.

This survey is of course not exhaustive. It nonetheless allows us to highlight certain points. First of all, making digital documents freely available takes a long time, and sources that could legally be accessible to the general public only become so several months or years later. This is understandable, of course, as digitization and online availability by cultural heritage institutions necessarily takes time. But this could get in the way of certain applications requiring the analysis of documents in "real time". Adapting the process of archiving could be necessary: providing natively digital document straight to an electronic archive harvestable by researchers seems the only feasible option. This is already the case for documents such as the United Kingdom parliamentary debates, through the online Hansard [15], for French laws and parliamentary debates with Legifrance [16]. Why not imagining a similar system for the international institutions? That kind of electronic archive could also help to solve problems of secrecy: some of the less sensitive documents, officially not open to the public, could be treated directly through the platform, in the same way as copyrighted texts on Hathitrust [17]: you cannot see the texts, but you can compute and export aggregates related to them. When are the moments in recent years when a specific topic has been the most discussed in closed negotiations? Even if I cannot access the transcripts of negotiations themselves, I could be able to get the information through the platform. Table 1.

The other possibility, much more costly, is of course that researchers would handle themselves the digitization and automated transcription of the relevant documents. A handful of scholarly initiatives have helped to make other sources available. More than 7000 country statements from the General Debate at the United Nations have for instance recently been digitized and analyzed [18], and fostered the development and testing of dedicated NLP tools [19]. A collaboration between different American universities helped build the Freedom of Information Archive Database, gathering around 3 million documents, mostly regarding the

Table 1

Diplomatical texts: examples of digital sources made available by official institutions.

COUNTRY	FORMAT	PDF and automated OCR, not always reliable
CANADA	Searchable PDF	PDF and automated OCR, not always reliable
FRANCE	Scans and PDF documents. Automated OCR provided, but hardly reliable	PDF and automated OCR, not always reliable
SWITZERLAND	Selection of archives transcribed and edited in collections (PDF easily OCREd). Other sources in raw PDF	PDF and automated OCR, not always reliable
UNITED NATIONS (UN)	Searchable scans and PDF documents	Official Document System
WORLD HEALTH ORGANIZATION (WHO)	PDF and automated OCR, not always reliable	Institutional Repository for Information Sharing (IRIS)

United States diplomacy during the XXth century - some documents dating back to as early as the XVIIIth century, the most recent dating from 2013 [20]. Regardless of these incredibly helpful initiatives, the amount of work to be done remains gigantic, and the lack of availability of the most recent documents prevents computational diplomacy to provide actual hands-on use cases using those texts.

2.2. The "new international relations" and the need to see beyond institutions

The context of the "new international relations" has made the debates surrounding any particular dispute more complex than ever. The resurgence in recent decades of bilateral negotiations in parallel with multilateral negotiations, the multiplication of actors involved (international institutions, NGOs, chambers of commerce etc.) and the development of corporate diplomacy make it more and more difficult to analyze each debate as a whole. Readiness thus means being able to analyze data coming from a wide array of institutions. Harvesting the web in search for opinions in the public, political leaders etc. is a process widely known and used elsewhere in the computational social sciences. But to reach actual readiness, we would need to extend the same processes as for international organizations to relevant NGOs, municipalities etc. involved in actual diplomatic actions.

3. How ready are our tools ?

Analyzing most of the data we need to handle in computational diplomacy compares to what is already done elsewhere in computational social sciences (Natural Language Processing, network analysis, predictive algorithms etc.). But even regarding the specific challenges faced by computational analysis such as adversarial behavior, we most likely seem to be ready.

Assessing the authenticity of certain sensitive documents will mostly be a task best performed of the intelligence services advising each state. But these private assessments are not necessarily sincerely shared (nor shared at all) with the relevant partners. Most importantly, they might not even have been performed at all, some misdirections or false documents being provided even in contexts task possibly considered of relatively minor importance. Purely scholarly projects will thus need to helping practitioners, or even more transparent and internationally-led intelligence initiative such as the UN Peacekeeping initiative [21], to assess the quality of the data they are provided with could be task well fit for the field of computational diplomacy. And some reliable and rather fast running tools already exist.

3.1. Authorship attribution

Authorship attribution or stylometry is meant to identify the author of a text by studying its linguistic properties. While it has found many applications in the humanities [22–24], it recently went on to be used in courts of justice [25,26], or in journalistic investigations [27,28]. The computational complexity of the task is rarely a problem, but the interpretation of the results can take some time: most AI-based analyses rely on easily interpretable methods [29], to be sure that the attribution is made on a set linguistically relevant features, rather than on topics evoked in the documents, typographic variations etc.

3.2. Detecting deepfakes

A growing threat for international safety, as recent examples during the war in Ukraine have shown [30], deepfakes could be detected [31] through procedures recently proposed to protect World Leaders [32].

3.3. Detecting fake news and profiling

Beyond the authenticity of a document, the quality of his content of matters a lot for the evaluation of a decision. Fake news detection algorithms could thus be of some use, especially as this recent field has very rapidly progressed. Recent benchmarks now show that an accuracy above 95% is to be expected [33]. The ever-changing nature of misinformation however makes the fight for accurate detection a never-ending fight [34]. More generally, profiling documents [35] could help understand the type of documents we are provided: has it been written by a corporation or an association? by a single individual or a collective? by a young student or by the person who signed the document? The performance vary depending on the complexity of the task, but F1-score are also regularly past the 95%.

4. Conclusion

Sophisticated data analysis techniques are at our disposal to ensure the quality of diplomacy-related data, and to make sense of complex and rapidly changing situations. But the main challenge remains to collect and properly disseminating data. Creating an infrastructure helping to respect the secrecy of certain documents while being able to compute on them, where institutions would routinely deposit their archives in electronic format would certainly be a cost-efficient and operational solution to a series of otherwise complex problems. Only then will we be able to design research protocols helpful for diplomatic decision-making.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

No data was used for the research described in the article.

Acknowledgements

I am thankful to the participants of the workshop "From the modeling of social behavior to computational diplomacy" organized by the Université de Genève and ETH Zürich, and especially to Thomas Chadeaux, Stephan Davidshofer, Jean-Luc Falcone and Nicolas Levrat for their helpful comments, as well as participants to the Human rights week conference at the Université de Genève and to the Ecole Nationale des Chartes Digital Humanities seminar. I am also grateful to François Falconet and Marie Puren for their useful remarks or advice. Errors

remain my own.

References

- [1] D.M. Berry, The computational turn: thinking about the digital humanities, *Cult. Mach.* 12 (2011).
- [2] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, *Computational social science*, *Science* 323 (2009) 721–723.
- [3] H.A. Ünver, Computational international relations what can programming, coding and internet research do for the discipline? *All Azimuth: A J. Foreign Policy Peace* 8 (2019) 157–182.
- [4] V. Pouliot, J. Cornut, *Practice theory and the study of diplomacy: a research agenda*, *Coop. Confl.* 50 (2015) 297–315.
- [5] D. Hart, A. Siniver, The meaning of diplomacy, *Int. Negot.* 26 (2020) 159–183.
- [6] F. Moretti, *Graphs, maps, trees: abstract models for a literary history*, Verso, 2005.
- [7] M. Kestemont, F. Karsdorp, E. de Bruijn, M. Driscoll, K.A. Kapitan, P.O. Macháin, D. Sawyer, R. Sleiderink, A. Chao, *Forgotten books: The application of unseen species models to the survival of culture*, *Science* 375 (2022) 765–769.
- [8] M. of Armed Forces AI Task Force, 2019. Artificial intelligence in support of defence, Paris, France: French Ministry of Armed Forces (2019).
- [9] P. Svenmarck, L. Luotinen, M. Nilsson, J. Schubert, Possibilities and challenges for artificial intelligence in military applications, in: *Proceedings of the NATO Big Data and Artificial Intelligence for Military Decision Making Specialists' Meeting*, pp. 1–16.
- [10] S.J. Bickley, H.F. Chan, B. Torgler, *Artificial intelligence in the field of economics*, *Scientometrics* 127 (2022) 2055–2084.
- [11] U. Nations, 1961. Vienna Convention on Diplomatic Relations, 1961.
- [12] R.D. Putnam, *Diplomacy and domestic politics: the logic of two-level games*, *Int. Organ.* 42 (1988) 427–460.
- [13] M. Puren, A. Pellet, N. Bourgeois, P. Vernus, F. Lebreton, *Between History and Natural Language Processing: Study, Enrichment and Online Publication of French Parliamentary Debates of the Early Third Republic (1881–1899)*, in: *ParlaCLARIN III at LREC2022-Workshop on Creating, Enriching and Using Parliamentary Corpora*.
- [14] B.C.G. Lee, *Machine learning, template matching, and the International Tracing Service digital archive: automating the retrieval of death certificate reference cards from 40 million document scans*, *Digit. Scholarsh. Humanit.* 34 (2019) 513–535.
- [15] D. Archer, *Negotiating difference in political contexts: an exploration of Hansard*, *Lang. Sci.* 68 (2018) 22–41.
- [16] S. Cottin, 2008. *Apprendre à décrire le document juridique à l'heure du Web 2.0*, *Lidil. Revue de linguistique et de didactique des langues* (2008) 53–70.
- [17] H. Christenson, *HathiTrust, Library Resources & Tech. Serv.* 55 (2011) 93–102.
- [18] A. Baturu, N. Dasandi, S.J. Mikhaylov, *Understanding state preferences with text as data: Introducing the UN General Debate corpus*, *Res. Polit.* 4 (2017), 2053168017712821.
- [19] K. Watanabe, Y. Zhou, *Theory-driven analysis of large corpora: Semisupervised topic classification of the UN speeches*, *Soc. Sci. Comput. Rev.* 40 (2022) 346–366.
- [20] M.J. Connelly, R. Hicks, R. Jervis, A. Spirling, C.H. Suong, *Diplomatic documents data for international relations: the Freedom of Information Archive Database*, *Confl. Manag. Peace Sci.* 38 (2021) 762–781.
- [21] S.-M. Martin-Brûlé, *Competing for trust: challenges in United Nations Peacekeeping-Intelligence*, *Int. J. Intell. Count.* 34 (2021) 494–524.
- [22] M. Kestemont, J. Stover, M. Koppel, F. Karsdorp, W. Daelemans, *Authenticating the writings of Julius Caesar*, *Expert Syst. Appl.* 63 (2016) 86–96.
- [23] P. Plechá, *Relative contributions of Shakespeare and Fletcher in Henry VIII: an analysis based on most frequent words and most frequent rhythmic patterns*, *Digit. Scholarsh. Humanit.* 36 (2021) 430–438.
- [24] F. Cafiero, J.-B. Camps, *Why Molière most likely did write his plays*, *Sci. Adv.* 5 (2019) eaax5489.
- [25] P. Juola, *Stylometry and immigration: a case study*, *J. Law Policy* 21 (2013).
- [26] F. Cafiero, J.-B. Camps, 2022. *Affaires de style*, Le Robert, 2022.
- [27] D.D. Kirkpatrick, 2022. *Who Is Behind QAnon? Linguistic Detectives Find Fingerprints*, *The New York Times* 19 (2022).
- [28] F. Cafiero, J.-B. Camps, *Who could be behind QAnon? Author Attrib. Supervised Mach. -Learn.* (2023).
- [29] M. Kestemont, E. Stamatatos, E. Manjavacas, W. Daelemans, M. Potthast, B. Stein, 2019. Overview of the cross-domain authorship attribution task at \$PPAN\$ 2019, in: *Working Notes of CLEF 2019: Conference and Labs of the Evaluation Forum*, Lugano, Switzerland, September 9–12, 2019, pp. 1–15.
- [30] B. Allyn, *Deepfake video of Zelensky could be 'tip of the iceberg' in info war, experts warn*, NPR, NPR, March 17 (2022).
- [31] M. Groh, Z. Epstein, C. Firestone, R. Picard, *Deepfake detection by human crowds, machines, and machine-informed crowds*, *Proc. Natl. Acad. Sci.* 119 (2022), e2110013119.
- [32] M. Bohá, H. Farid, *Protecting world leaders against deep fakes using facial, gestural, and vocal mannerisms*, *Proc. Natl. Acad. Sci.* 119 (2022), e2216035119.

- [33] G. Gravanis, A. Vakali, K. Diamantaras, P. Karadais, Behind the cues: a benchmarking study for fake news detection, *Expert Syst. Appl.* 128 (2019) 201–213.
- [34] S.I. Manzoor, J. Singla, 2019. Fake news detection using machine learning approaches: A systematic review, in: 2019 3rd international conference on trends in electronics and informatics (ICOEI), pp. 230–234.
- [35] G.K. Mikros, Blended Authorship , 2017. Attribution: Unmasking Elena Ferrante Combining Different Author Profiling Methods, *UPPADO* (2017) 85.



Florian Cafiero is a statistician at the Medialab of the Institut d'Études Politiques de Paris and a lecturer in digital humanities at the École nationale de chartes - PSL. He has taught at the École centrale de Paris, the University of Geneva and the University of Paris-Sorbonne. In 2022 he is a visiting scholar at Columbia University. His work in digital humanities and computational social sciences has been published in journals such as *Science Advances*, *Social Networks* or *Social Science and Medicine*, and presented in conferences at Columbia, ETH Zürich, Harvard, MIT, NYU, Rice, la Sapienza, UCLA etc. With Jean-Baptiste Camps, he is the author of a book on stylometry, *Affaires de style*, published by Le Robert. For an investigation by Pulitzer Prize winner David Kirkpatrick, the two of them co-authored a study on the potential authors hiding behind the Q of QAnon published in the *New York Times*. Since 2019, he has collaborated with the Conseil Scientifique de l'Éducation Nationale, as a member of the "Developing Critical Thinking" working group and then joined the work of the "Equity and Reducing Educational Inequality" working group. He is also the co-author of a report on "sectarian drifts on the internet" for the Interministerial Mission of Vigilance and Combat against Sectarian Drifts of the French Ministry of the Interior (MIVILUDES), and of a report on racism online commissioned by the National Consultative Commission on Human Rights of the French Ministry of Foreign Affairs. Florian Cafiero is an alumnus of the École normale supérieure Paris-Saclay a graduate of the École nationale des chartes.