ICC
Dispute Resolution
Services

ICC COMMISSION DOCUMENT

The use of AI to assist in the identification, validation and assessment of red flags of corruption

Offprint from ICC Dispute Resolution Bulletin | ISSUE 2026-2

ICC Dispute Resolution Bulletin | 2026-2
Addressing Issues of Corruption in International Arbitration

E-journal of the International Chamber of Commerce (ICC)
Périodique numérique de la Chambre de commerce internationale

Acknowledgements
The ICC Commission on Arbitration and ADR Task Force "Addressing Issues of Corruption in International Arbitration" was co-chaired by **Sophie Nappert**, **José Ricardo Feris** and **Vladimir Khvalei**, under the leadership of **Carita Wallgren-Lindholm**, previous Chair of the ICC Commission on Arbitration and ADR and **Melanie van Leeuwen**, current Chair of the ICC Commission on Arbitration and ADR, **Patrick Leonard** and **Caline Mouawad**, Steering Committee Liaison Officers, and **Dr Hélène van Lith**, Steering Committee Officer and Commission Secretary. The Task Force has worked closely with the International Bar Association (IBA), including its arbitration and anti-corruption commissions, represented by Julie Bédard and Dietmar Prager, as well as with the ICC Business Integrity Commission, led by Viviane Schiavi.

Track coordinators

Red flags or other indicators of corruption
Xavier Andrade Cadena, Patrick Baeten, Lucinda A. Low, Abdulhay Sayed

National committee survey
Diana Bayzakova, Payel Chatterjee, Ana Gerdau De Borja Mercereau, Christine Lecuyer-Thieffry

The use of AI to assist in the identification, validation and assessment of red flags of corruption
Niuscha Bassiri, Polina Permyakova, Reinmar Wolff.
Experts: Battine Edwards, Henry Liu, Lilit Nagapetyan

Burden and standard of proof of corruption
Laurie Achtouk-Spivak, Arif Hyder Ali, Françoise Lefèvre, Cédric de Pouzilhac

Issues related to arbitrators
Clare Ambrose, Michael Hwang, Patricia Nacimiento.

Parallel Corruption Proceedings
Paola Aldrete Rivas, Elena Gutiérrez García De Cortázar, Susan Munro

Other Task Force members: The Task Force was launched on 5 November 2019, and its final membership comprised members from 58 jurisdictions: Diana Akikol, Folashade Alli, Koorosh H. Ameli, Ricardo Ampuero Llerena, Nicolas Angelet, Homayoon Arfazadeh, Andrea Bandini, Victoria Barth, Antje Baumann, Julie Bédard, Affef Ben Mansour, Sílvia Bueno De Miranda, Constantin Calavros, Michelangelo Cicogna, Massimo Coccia, Nadia Darwazeh, Sotiris Dempegiotis, Diana Droulers, Nils Eliasson, Alain Farhad, Marcelo Carlos H. Fernandez Flores, Cláudio Finkelstein, Jun (Gary) Gao, Brody K. Greenwald, Jacob Grierson, Renato Stephan Grion, Florian Haugeneder, Wulf Gordian Hauser, Emily Hay, Roberto Hernández García, Faith Hochberg, Mélida Hodgson, Junichi (Tim) Horakova, Lanlan Hu, Mahmood Hussain, Therese Isaksson, Talal H. Jaber, Sameer Jain, John Kadelburger, Hussein Khaddour, Christian Konrad, Philipp Kotlaba, Bartosz Kruzewski, Iohann Le Frapper, Pierrick Le Goff, Zhuo Jun Jennifer Lim, Aloysius (Louie) Llamzon, Esteban López, Eliseo Martínez, Hugh Meighen, Ian Meredith, Juan Felipe Merizalde Urdaneta, Catarina Monteiro Pires, Renato Nazzini, Piotr Nowaczyk, Babajide Ogundipe, Oluwaseyilayo Ojo, Julieta Ovalle Piedra, Marily Paralika, Olena Perepelynska, José María Pérez, Victor Radnaev, Peter Rižnik, Silvia Rodríguez Vasquez, Anita Schieffer, Daniel Schimmel, Thomas Stouten, Marc Thill, Filipe Vaz Pinto, Ana Vermal, Carlo Verona, María Vicien Milburn, Sophia Von Dewall, Stephan Wilske, Katia Yannaca-Small, Hua Hsin Yen, Jennifer Younan, Haixiao Hellen Zhang, Roland Ziadé, and Alberto Zuleta-Londoño.

Copyright © 2026
International Chamber of Commerce (ICC)

All rights reserved. No part of this work may be reproduced, distributed, transmitted, translated or adapted in any form or by any means except as permitted by law without the written permission of ICC. Authors should request the permission of ICC prior to any reproduction or subsequent publication of an article (electronic or print). Permission can be requested from ICC through **ipmanagement.drs@iccwbo.org**.

Tous droits réservés. Il est interdit de reproduire, de distribuer, de transmettre, de traduire ou d'adapter tout ou partie de cet ouvrage, sous quelque forme ou par quelque procédé que ce soit, en dehors des exceptions prévues par la loi, sans l'autorisation écrite de la Chambre de commerce internationale. Toute demande d'autorisation est à adresser à **ipmanagement@iccwbo.org**.

Disclaimer
Except where otherwise indicated, the views expressed and statements made herein are those of their authors and should not be construed as creating any duty, liability or obligation on the part of ICC and its constituent bodies, including the International Court of Arbitration, the International Centre for ADR and their respective Secretariats.

Sauf indication contraire, les points de vue et les commentaires exprimés dans la présente publication sont ceux de leur(s) auteur(s) et ne sauraient créer aucun devoir, ni aucune responsabilité ou obligation à la charge de la Chambre de commerce internationale ou de ses organes, y compris la Cour internationale d'arbitrage, le Centre international d'ADR et leurs secrétariats respectifs.

Trademarks

ICC, the ICC logo, CCI, International Chamber of Commerce (including Spanish, French, Portuguese and Chinese translations), International Court of Arbitration and ICC International Court of Arbitration (including Spanish, French, German, Arabic and Portuguese translations) are all trademarks of the ICC, registered in several countries.

ICC, le logo ICC, CCI, International Chamber of Commerce (y compris des traductions en espagnol, français, portugais et chinois), International Court of Arbitration et ICC International Court of Arbitration (y compris des traductions en espagnol, français, allemand, arabe et portugais) sont des marques de la Chambre de commerce internationale et ont été enregistrées dans plusieurs pays.

ICC Publication No. @26BUL2
ISBN: 978-92-842-0773-2 ISSN: 2520-6052
publications@iccwbo.org

https://2go.iccwbo.org/

ICC Dispute Resolution Library
https://jusmundi.com/en/icc-dispute-resolution-library

Publication date | Date de parution
November 2026 | Novembre 2026

Contents

THE USE OF AI TO ASSIST IN THE IDENTIFICATION, VALIDATION AND ASSESSMENT OF RED FLAGS OF CORRUPTION

> **Introduction** 4

> **1. Potential uses of AI in arbitrations involving allegations of corruption** 4

> **2. Takeaways to bear in mind when considering the use of AI to identify, validate and assess corruption red flags in arbitration** 6

2.1 Choice of AI solutions 6

2.2 Probative value of AI-generated outputs 6

2.3 Data availability 6

2.4 Efficiency and expeditiousness, and cost-benefit considerations 6

> **Conclusion** 6

The Use of AI to Assist in the Identification, Validation and Assessment of Red Flags of Corruption

Introduction

1. AI is playing an increasingly important role as an effective and necessary tool to support searches for indicia of corruption in compliance and internal investigation processes. Data analytics tools with AI elements can efficiently review very large datasets, find associations and patterns that are not obvious to the human mind, and improve the identification, monitoring and response to higher risk transactions and relationships.

2. In an arbitration involving allegations of corruption, parties and tribunals often need to consider whether a business transaction or a relationship is tainted by or involves incidents of corruption, commonly in the absence of any direct evidence of corruption. Identifying possible direct and/or circumstantial evidence of corruption requires significant time, effort and financial resources, particularly in an environment where rogue actors can be very sophisticated.

3. This Supplementary Document ("Document") explores the degree to which currently available AI tools can assist in the identification, validation and assessment of red flags in the arbitration context.[1]

4. While this Document does *not* explore the extent to which AI tools can be made compatible with the procedural and substantive tenets of the arbitration process, it focuses on the exploration of what AI tools can and cannot do, from a technical perspective, to assist arbitration practitioners in identifying, validating and assessing red flags, on the assumption that those AI tools are properly and safely integrated in the arbitration process.[2]

1. Potential uses of AI in arbitrations involving allegations of corruption

5. The use of AI for identifying and analysing red flags in arbitration is in its nascent stages.

6. As was insightfully explained and predicted by the technical experts sessions, AI-powered document review tools offer automated assistance in reviewing data and identifying potential indicators of corruption in such data.[3] This assistance is based on manual instructions on the types of red flags to be identified.

7. However, the tasks of critical validation and assessment of red flags in the circumstances of a case are not (yet) addressed by commercially available AI-powered tools. They might be addressed in the medium term, once significant logical mapping is carried out and the necessary resources are available.

8. Furthermore, there is no indication on current information to suggest that AI-generated outputs have been relied upon as corruption-related evidence in arbitration proceedings, nor is there an indication that custom-made AI-powered algorithms have been devised to assist with the detection of corruption in a case.

9. Nevertheless, it is conceivable that, with the important advancements in AI technology and its developing regulation, parties and their expert(s), as well as tribunals and tribunal-appointed experts, may consider making use of AI to aid in the identification[4] and validation[5] of red flags and the assessment[6] of the likelihood of corruption.

---

1 These three terms ("identification", "validation" and "assessment" are in reference to the ICC Commission Supplementary Document *Red Flags or Other Indicators of Corruption in International Arbitration*.

2 These and other related questions are currently being addressed as part of ongoing work of the ICC Commission Task Force on Artificial Intelligence in Dispute Resolution.

3 See e.g. A. Petheram, I. Nti Asare, Riom Open Data to Artificial Intelligence: the Next Frontier in Anti-Corruption (https://oxfordinsights.com, 27 July 2018); and N. Köbis, C. Starke, I. Rahwan, Artificial Intelligence as an Anti-Corruption Tool (AI-ACT) – Potential and Pitfalls for Top-down and Bottom-up Approaches (Feb. 2021).

4 The identification of red flags of corruption refers to an issue-spotting exercise in relation to facts, factors, or circumstances of relevance to the specific corrupt practice at issue (either as alleged or as appear) that indicate the potential risk of that corrupt practice (or not) such that they should be investigated further. See *Red Flags or Other Indicators of Corruption in international Arbitration*, section 2.1 "A three-step methodology".

5 The validation of red flags refers to an exercise aimed at critically confirming (or negating) identified red flags, including by verifying whether they are factually supported and by verifying their strength. See *Red Flags or Other Indicators of Corruption in international Arbitration*, section 2.1 "A three-step methodology".

6 The assessment of red flags refers to an exercise aimed at assembling and evaluating the composite evidentiary picture that emerges from the totality of the relevant facts and circumstances, including consideration of red flags, contrary indicators or "green" flags, and other mitigating measures or circumstances. See *Red Flags or Other Indicators of Corruption in international Arbitration*, section 2.1 "A three-step methodology".

10. Particularly in cases involving large amounts of data, AI can search efficiently for both general and specific red flags, uncover patterns and associations that escape the human mind, and identify missing data. Drawing from big data, AI tools also have the potential to be more consistent and more accurate than a human decision-maker when validating and assessing potential corruption.

11. More specifically, demos presented by Deloitte to present the specific workings of its AI tool *NavigAIte*[7] showed that AI-powered document review tools are particularly useful in identifying given red flags in a dataset, also in cases involving large volumes of data in different languages.

12. Impressively, training can enable algorithmic sentiment recognition, whereby the AI tool can not only interpret what is communicated, but also how statements are communicated through sentiment analysis, indicating human emotions such as enthusiasm, anger, impatience or trepidation.[8] In this manner, one can foresee that an AI tool could detect the intent of potential fraudulent conduct.[9]

13. That said, the identification of red flags currently requires manual human input in the form of instructions on the types of red flags that the tool should look for, e.g. country risk, transaction features and excessive payments.[10] On the latter, though, it is remarkable that the (properly-trained) AI tool is able to detect an excessive payment without any instruction as to what would be a generally acceptable payment amount.[11]

14. The AI technology in the document review tools is not currently designed to provide opinions, including opinions on the likelihood of corruption. An important collection of awards and decisions on corruption and fraud was assembled to train *NavigAite* for demonstrative purposes with factual circumstances, considerations, reasonings and determinations on red flags, as well as the ICC Supplementary Document *Red Flags or Other Indicators of Corruption in International Arbitration*. The current mapping of *NavigAite*, however, requires further enhancement to generate output on the likelihood of corruption in a dataset.

15. While AI tools are being used by internal audit sampling systems and accountants to spot issues for investigation, public authorities also make use of AI to assist in social benefits fraud investigations (albeit they have attracted criticism for the use of allegedly discriminatory algorithms, or even discontinued by court orders).[12]

16. At this early stage of regulation, legal rules and policies do not prohibit the use of AI for any of the mentioned purposes but may provide for certain heightened obligations affecting its use, amongst them the obligation to disclose the use of AI tools.[13]

17. In all events, the importance of data quality and availability remains. As arbitration participants may not have access to the complete dataset in a corruption case, conceptual and contextual analysis using AI is more difficult. This may result in a less-than-optimal reliability of the output. Given the likelihood of requiring a multi-source analysis involving gaming, chats, mobile devices, call recordings, transaction logs and other statements, in addition to email correspondence, to confirm a potential red flag, obtaining access to these data sources could be a significant challenge to using AI in arbitration.

---

7 https://www.deloitte.com/nl/en/services/consulting-financial/services/deloitte-navigaite.html

8 See also Z. Elyoseph et al., ChatGPT Outperforms Human in Emotional Awareness Evaluations, *Frontiers in Psychology* 14 (2023). ChatGPT's summary of this article states as follows: "This study investigates ChatGPT's ability to identify and describe emotions using the Levels of Emotional Awareness Scale (LEAS). The findings suggest that ChatGPT can generate appropriate emotional awareness responses, with performance that may improve significantly over time. The authors discuss potential theoretical and clinical implications, including the use of ChatGPT in cognitive training for individuals with emotional awareness impairments and its application in psychiatric diagnosis and assessment."

9 See e.g. Ketrone - Legal AI.

10 On typologies and common types of red flags, see *Red Flags or Other Indicators of Corruption in international Arbitration*, section 1(5).

11 For example, in a test run with a specifically prepared dataset for the research supporting this Document, *NavigAite* detected that the sales price for cabin crew outfits must have been substantially lower than the sales price for an aircraft. It detected the document containing the inflated sales price offer for the cabin crew outfits as a potential red flag, without any specific human input on the appropriate pricing of cabin crew outfits.

12 See, e.g. Sweden: Authorities must discontinue discriminatory AI systems used by welfare agency (www.amnesty.org, 24 Nov. 2024) SyRI (System Risk Indication) case, District Court of The Hague, *Rechtbank Den Haag* - docket number C-09/550982 HA ZA 18/338.

13 For example, pursuant to the Regulation (EU) 2024/1689 (Artificial Intelligence Act), deployers of AI may be required to take appropriate technical and organizational measures to ensure that AI systems are used in accordance with their instructions, assign human oversight to natural persons who have the necessary competence, training, authority and support, monitor the operation of deployed AI systems, ensure that the input data is relevant and sufficiently representative to the extent they exercise control over it, keep the logs automatically generated by the system to the extent such logs are under their control, and carry out data protection assessments. See Art. 6(2), Annex III, para. (8), and Art. 26 of the Regulation (EU) 2024/1689. M. Scherer, O. Jensen, R. Childree, "Regulating the Use of Artificial Intelligence in International Arbitration: The EU AI Act and Beyond", *The Paris Journal of International of Arbitration*, 2024-3, p. 653; see also, e.g. the judgment of Hon. Jonathan G. Schopf, Matter of Weber 2024 NY Slip Op 24258 on 10 Oct. 2024, which provides insights in the types of evidentiary challenges which may arise in the United States due to the lack of transparency, explainability and record keeping in expert evidence and some recent guidance in relation to the same issued by the Expert Witness Institute in the United Kingdom, How not to use AI in expert evidence (www.ewi.org.uk, 25 Oct. 2024).

18. Another overarching issue is the confidentiality aspect from a technical perspective, as the arbitration dataset will be uploaded to a "third party" to the arbitration. Delegates, in particular in-house counsel, have expressed the concern of maintaining data confidential within the AI tool. While systems like *NavigAIte* use a closed system, the concern remains that the information provided to the AI tool manager – being human – might be disclosed to yet another set of third parties. However, this concern also applies to other participants in an arbitration proceeding, such as institutions and court reporters, translators and interpreters, for which safeguards have been successfully set up.

2. Takeaways to bear in mind when considering the use of AI to identify, validate and assess corruption red flags in arbitration

2.1 Choice of AI solutions

19. Criteria and features to be considered when choosing an AI solution include:

- Transparency, explainability/rationale and verifiability of the output of a specific AI solution to meet due process, evidentiary and award enforcement requirements;[14]
- Confidentiality of data, including whether the tool uses input data for training and otherwise ensures the confidentiality, by being typically designed as closed systems, without an online version; and
- Data protection features, to meet data protection requirements, in particular when personal data is uploaded for training or evaluation purposes.

2.2 Probative value of AI-generated outputs

20. Limitations of AI/AI-generated outputs that need to be taken into account when considering the probative value of AI-generated outputs could stem from potential flaws and biases embedded in the algorithm and the training data used for its design and from a lack of a specific standard of proof embedded in an AI solution.[15]

21. Flaws and biases may result from, among others:

- Class imbalances (i.e. a lack of data diversity in a training set: if the dataset includes very few instances of corruption-related transactions – corporate payments of bribes typically take place infrequently at irregular intervals – the model can conclude that bribes never happen);
- Overfitting (where the model cannot generalise or fit well on a dataset other than the training dataset); and
- Underfitting (where the model does not work well with either the training dataset or the test dataset).

2.3 Data availability

22. The use of existing commercial generative AI tools does not involve any additional training of the tool (for which large amounts of training data is required) and can review any dataset, even small datasets. However, data availability may become an issue where an AI tool is being developed for a particular case, as the training data is further limited to one party's side only, or, at best of the entire arbitration file.

23. Voluntary or ordered document or, rather, data production will play a key role in arbitration proceedings for the purposes of training an AI tool, which would encompass a multi-source analysis.

2.4 Efficiency and expeditiousness, and cost-benefit considerations

24. The existing AI technology does not raise any distinct issues impacting the efficiency and expeditiousness of proceedings. The tribunal's use of AI on its own initiative can be compared to the use of a tribunal-appointed expert.

Conclusion

25. While, at this moment, the use of AI to assist with the identification, validation and assessment of red flags in arbitration may be limited to the first component of identification, on the basis of the available AI technology and anticipating its further rapid development, it would be premature to exclude the ability of AI tools to validate and assess red flags. Whether and to what extent legal impediments arise to the use of AI in any context in arbitration is thoroughly being considered by the ongoing ICC Commission Task Force "Artificial Intelligence in Dispute Resolution".

14 See e.g. Recital (9) of the Regulation (EU) 2024/1689 (Artificial Intelligence Act).

15 As regards flaws and biases, see Coalition for Integrity's paper Using machine learning for anti-corruption risk and compliance (2021). See also OECD documents: Due Diligence Guidance for Responsible Business Conduct (2018); Analytics for Integrity. Data-Driven Approaches for Enhancing Corruption and Fraud Risk Assessment (2019) and Anti-Corruption and Integrity Outlook (2026).