

# John Pavlopoulos

 ipavlopoulos |  ipavlopoulos |  ipavlopoulos.github.io |  ipavlopoulos@aueb.gr |

## SUMMARY

---

Assistant Professor specializing in Machine Learning for Information Retrieval and Data Mining of large document collections at the Department of Informatics, Athens University of Economics and Business, Greece. He holds affiliations with Archimedes/Athena RC, Greece, and Stockholm University, Sweden. Previously, he served as a Senior Lecturer (fixed-term) in NLP and Data Science at the Department of Computer and Systems Sciences, Stockholm University, Sweden, and was a visiting scholar at the Venice Centre for Digital and Public Humanities, Ca' Foscari University of Venice, Italy. His research focuses on Machine Learning for Natural Language Processing (NLP), with a strong emphasis on applications in Digital Humanities. His work has been published in top-tier journals, such as *Machine Learning* and *Nature*, and presented at leading NLP conferences, such as ACL and EMNLP.

## WORK EXPERIENCE

---

<b>Assistant Professor</b>	May 2024 - present
Department of Informatics, Athens University of Economics and Business, Greece <i>Machine Learning for Information Retrieval and Data Mining of Large Document Collections</i>	
<b>Visiting Scholar</b>	Dec 2021 - Aug 2022
The Venice Centre for Digital and Public Humanities (VeDPH), Ca'Foscari, Italy	
<b>Senior Lecturer (fixed-term)</b>	Jun 2020 - Nov 2021
Department of Computer and System Sciences, Stockholm University, Sweden	
<b>Postdoctoral researcher</b>	Jun 2019 - May 2020
Department of Computer and System Sciences, Stockholm University, Sweden	

## PROJECTS

---

<b>LLM3</b>	<a href="#">Link to Website</a>
This project develops and evaluates mediation agents based on Large Language Models that will actively participate in online discussions, with or without additional human mediation.	
<b>EFRA</b>	<a href="#">Link to Website</a>
EU-funded project (2023-2026) for AI-enabled food risk prevention.	
<b>NLP for Digital Art History</b>	<a href="#">Link to Website</a>
A combination ‘distant viewing’ of visual materials with NLP and ‘close reading’ of the artefacts through formal and contextual analysis to study Japanese early modern woodblock prints, so-called ukiyo-e (‘pictures of the floating world’), produced between the seventeenth and mid-nineteenth century.	

## EDUCATION

---

2010 - 2014	PhD (Aspect-based Sent. Analysis) at <b>Athens University of Economics and Business</b>
2008 - 2009	MSc (AI) at <b>University of Edinburgh</b>
2001 - 2007	BSc (Applied Math. & Phys. Sciences) at <b>National Technical University of Athens</b>

## GRANTS

---

- 2025-2026 Funding for junior faculty “Action 1” (AUEB).
- 2022-2023 Studying toxic language classifiers (unrestricted research support by Google).
- 2020-2021 Cost model development for the moderation of user posts on online platforms (project with the Joint Research Center, European Commission; PI).
- 2019-2020 Context-aware toxic language detection in online conversations (Google Research Award; PI).
- 2016-2018 Machine learning to filter profanity (Google Digital News Initiative; PI).

## PUBLICATIONS

---

- Assael, Yannis, Thea Sommerschield, Alison Cooley, et al. (2025). “Contextualizing ancient texts with generative neural networks”. In: *Nature* 645.8079, pp. 141–147.
- Bakagianni, Juli, Kalliopi V Dalakleidi, et al. (2025). “DiaShift: An Explainable System for Temporal Diagnostic Shift Detection in Clinical Notes”. In: *2025 IEEE 25th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE, pp. 341–345.
- Bakagianni, Juli, Kanella Pouli, et al. (2025). “A systematic survey of natural language processing for the Greek language”. In: *Patterns* 6.11.
- Bakagianni, Juli, Korbinian Randl, et al. (2025). “FoodSafeSum: Enabling Natural Language Processing Applications for Food Safety Document Summarization and Analysis”. In: *Conference on Empirical Methods in Natural Language Processing (EMNLP), November 2025, Suzhou, China*. Association for Computational Linguistics, pp. 16786–16804.
- Boumpi, Maria et al. (2025). “From Pen to Prediction: Handwriting-Based Alzheimer’s Detection”. In: *2025 IEEE 25th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE, pp. 671–675.
- Korre, Katerina, Dimitris Tsirmpas, et al. (2025). “Evaluation and facilitation of online discussions in the llm era: A survey”. In: *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 24454–24473.
- Randl, Korbinian et al. (2025). “Mind the gap: from plausible to valid self-explanations in large language models”. In: *Machine Learning* 114.10, p. 220.
- Barmpounis, Epameinondas-Konstantinos et al. (2024). “Recognition and Analysis of the Proceedings of the Greek Parliament after WWII.” In: *DHQ: Digital Humanities Quarterly* 18.1.
- De Gregorio, Giuseppe et al. (2024). “A New Framework for Error Analysis in Computational Paleographic Dating of Greek Papyri”. In: *International Conference on Document Analysis and Recognition*. Springer, pp. 102–118.
- Pavlopoulos, John, Maria Konstantinidou, Elpida Perdiki, et al. (2024). “Explainable dating of greek papyri images”. In: *Machine Learning* 113.9, pp. 6765–6786.
- Pavlopoulos, John and Aristidis Likas (2024). “Polarized Opinion Detection Improves the Detection of Toxic Language”. In: *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1946–1958.
- Pavlopoulos, John, Alv Romell, et al. (2024). “Automotive fault nowcasting with machine learning and natural language processing”. In: *Machine Learning* 113.2, pp. 843–861.

- Pavlopoulos, John, Ryan Sandell, et al. (2024). “HoLM: Analyzing the Linguistic Unexpectedness in Homeric Poetry”. In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 8166–8172.
- Pavlopoulos, John et al. (2024). “Revisiting silhouette aggregation”. In: *International Conference on Discovery Science*. Springer, pp. 354–368.
- Pavlopoulos, John et al. (2024). “Towards a Greek Proverb Atlas: Computational Spatial Exploration and Attribution of Greek Proverbs”. In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 11842–11854.
- Picca, Davide and John Pavlopoulos (2024). “Deciphering Emotional Landscapes in the Iliad: A Novel French-Annotated Dataset for Emotion Recognition”. In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 4462–4467.
- Toumazatos, Anastasios et al. (2024). “Still All Greeklish to Me: Greeklish to Greek Transliteration”. In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 15309–15319.
- Boulieris, Petros et al. (2023). “Fraud detection with natural language processing”. In: *Machine Learning*, pp. 1–22.
- Korre, Katerina, John Pavlopoulos, et al. (2023). “Harmful Language Datasets: An Assessment of Robustness”. In: *The 7th Workshop on Online Abuse and Harms (WOAH)*, pp. 221–230.
- Liagkou, Konstantina et al. (2023). “Leveraging the Spatiotemporal Analysis of Meisho-e Landscapes”. In: *International Conference on Discovery Science*. Springer, pp. 416–430.
- Nanetti, Andrea et al. (2023). “Sentiment Analysis of Primary Historical Sources”. In: *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, pp. 767–772.
- Pavlopoulos, John and Maria Konstantinidou (2023). “Computational authorship analysis of the Homeric poems”. In: *International Journal of Digital Humanities* 5.1, pp. 45–64.
- Pavlopoulos, John, Maria Konstantinidou, Isabelle Marthot-Santaniello, et al. (2023). “Dating Greek papyri with text regression”. In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 10001–10013.
- Pavlopoulos, John, Maria Konstantinidou, Georgios Vardakas, et al. (2023). “Explaining the Chronological Attribution of Greek Papyri Images”. In: *International Conference on Discovery Science*. Springer, pp. 401–415.
- Pavlopoulos, John, Vasiliki Kougia, Paraskevi Platanou, et al. (2023). “Detecting Erroneous Handwritten Byzantine Text Recognition”. In: *Findings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)*. Association for Computational Linguistics.
- Pavlopoulos, John and Aristidis Likas (2023). “Distance from Unimodality: Assessing Polarized Opinions in Abusive Language Detection”. In: *The 61st Annual Meeting Of The Association For Computational Linguistics*.
- Sommerschield, Thea et al. (2023). “Machine learning for ancient languages: A survey”. In: *Computational Linguistics* 49.3, pp. 703–747.

- Sorensen, Jeffrey et al. (2023). "JUAGE at SemEval-2023 Task 10: Parameter Efficient Classification". In: *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pp. 1195–1203.
- Assael, Yannis, Thea Sommerschield, Brendan Shillingford, et al. (2022). "Restoring and attributing ancient texts using deep neural networks". In: *Nature* 603.7900, pp. 280–283.
- Dritsa, Konstantina et al. (2022). "A greek parliament proceedings dataset for computational linguistics and political analysis". In: *Advances in Neural Information Processing Systems* 35, pp. 28874–28888.
- Korre, Katerina and John Pavlopoulos (2022). "Enriching grammatical error correction resources for Modern Greek". In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 4984–4991.
- Liagkou, Konstantina et al. (2022). "A study of distant viewing of ukiyo-e prints". In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 5879–5888.
- Paparigopoulou, Asimina et al. (2022). "Dating Greek papyri images with machine learning". In: *ICDAR Workshop on Computational Paleography*, <https://doi.org/10.21203/rs.3.rs-1000000>. Vol. 3.
- Pavlopoulos, John, Vasiliki Kougia, Ion Androutsopoulos, et al. (2022). "Diagnostic captioning: a survey". In: *Knowledge and Information Systems* 64.7, pp. 1691–1722.
- Pavlopoulos, John, Leo Laugier, et al. (2022). "From the detection of toxic spans in online discussions to the analysis of toxic-to-civil transfer". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 3721–3734.
- Pavlopoulos, John et al. (2022). "Sentiment analysis of Homeric text: The 1st Book of Iliad". In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 7071–7077.
- Platanou, Paraskevi et al. (2022). "Handwritten paleographic Greek text recognition: A century-based approach". In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 6585–6589.
- Xenos, Alexandros, John Pavlopoulos, Ion Androutsopoulos, et al. (2022). "Toxicity detection sensitive to conversational context". In: *First Monday*.
- Zonios, Christos et al. (2022). "Transformer-Based Music Language Modelling and Transcription". In: *Proceedings of the 12th Hellenic Conference on Artificial Intelligence*, pp. 1–8.
- Korre, Katerina et al. (2021). "ELERRANT: Automatic grammatical error type classification for Greek". In: *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pp. 708–717.
- Kougia, Vasiliki and John Pavlopoulos (2021). "Multimodal or text? retrieval or BERT? benchmarking classifiers for the shared task on hateful memes". In: *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*, pp. 220–225.
- Kougia, Vasiliki, John Pavlopoulos, Panagiotis Papapetrou, et al. (2021). "RTEX: A novel framework for ranking, tagging, and explanatory diagnostic captioning of radiography exams". In: *Journal of the American Medical Informatics Association* 28.8, pp. 1651–1659.
- Ljungman, Jimmy et al. (2021). "Automated grading of exam responses: an extensive classification benchmark". In: *Discovery Science: 24th International Conference, DS 2021, Halifax, NS, Canada, October 11–13, 2021, Proceedings* 24. Springer, pp. 3–18.

- Miliou, Ioanna et al. (2021). "Sentiment nowcasting during the COVID-19 pandemic". In: *Discovery Science: 24th International Conference, DS 2021, Halifax, NS, Canada, October 11–13, 2021, Proceedings* 24. Springer, pp. 218–228.
- Olczak, Jakub et al. (2021). "Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal". In: *Acta orthopaedica* 92.5, pp. 513–525.
- Pavlopoulos, John and Panagiotis Papapetrou (2021). "Customized Neural Predictive Medical Text: A Use-Case on Caregivers". In: *Artificial Intelligence in Medicine: 19th International Conference on Artificial Intelligence in Medicine, AIME 2021, Virtual Event, June 15–18, 2021, Proceedings*. Springer, pp. 438–443.
- Pavlopoulos, John, Jeffrey Sorensen, et al. (2021). "SemEval-2021 task 5: Toxic spans detection". In: *Proceedings of the 15th international workshop on semantic evaluation (SemEval-2021)*, pp. 59–69.
- Xenos, Alexandros et al. (2021). "Context sensitivity estimation in toxicity detection". In: *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*, pp. 140–145.
- Korre, Katerina and John Pavlopoulos (2020). "Errant: Assessing and improving grammatical error type classification". In: *Proceedings of the The 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pp. 85–89.
- Kougia, Vasiliki et al. (2020). "Medical image tagging by deep learning and retrieval". In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 11th International Conference of the CLEF Association, CLEF 2020, Thessaloniki, Greece, September 22–25, 2020, Proceedings 11*. Springer, pp. 154–166.
- Pavlopoulos, John and Panagiotis Papapetrou (2020). "Clinical predictive keyboard using statistical and neural language modeling". In: *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, pp. 293–296.
- Pavlopoulos, John et al. (2017). "Deeper attention to abusive user content moderation". In: *Proceedings of the 2017 conference on empirical methods in natural language processing*, pp. 1125–1135.
- Tsatsaronis, George et al. (2015). "An overview of the BIOASQ large-scale biomedical semantic indexing and question answering competition". In: *BMC bioinformatics* 16, pp. 1–28.
- Pavlopoulos, John and Ion Androutsopoulos (2014a). "Aspect term extraction for sentiment analysis: New datasets, new evaluation measures and an improved unsupervised method". In: *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)*, pp. 44–52.
- (2014b). "Multi-granular aspect aggregation in aspect-based sentiment analysis". In: *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 78–87.
- Pontiki, Maria et al. (2014). "SemEval-2014 Task 4: Aspect Based Sentiment Analysis". In: *SemEval 2014*, p. 27.
- Voudigari, Elli et al. (2011). "A framework for web Page Rank prediction". In: *International Conference on Engineering Applications of Neural Networks*. Springer, pp. 240–249.
- Pavlopoulos, John, John Vrettaros, et al. (2008). "The Development of a Self-assessment System for the Learners Answers with the Use of GPNN". In: *Emerging Technologies and Information Systems for the Knowledge Society: First World Summit on the Knowledge Society, WSKS 2008, Athens, Greece, September 24-26, 2008. Proceedings 1*. Springer, pp. 332–340.

Last updated: February 18, 2026