



Nikiforos Pittaras, PhD

Machine Learning researcher / engineer

Machine Learning, AI and Data Science enthusiast, with a passion for research and development of deep learning approaches for text and multimedia.

<https://npit.github.io>

pittarasnikif@gmail.com

`</>` python, java, matlab, C++
`>` sklearn, pytorch, tensorflow
`>` mysql, postgres, cassandra
`>` linux, bash, docker, git

EDUCATION

- **PhD in Machine Learning:** (DIT-UoA) 2017-2021
Joint industrial 4-year PhD programme ¹, Grade 'Excellent'.
Specialized in augmenting Deep Neural Representations with external information for various ML tasks.
Developed a Hate Speech Detection system from research findings for industry deployment
> python, tensorflow, keras, pytorch, DNN, nltk, Wordnet, ray, mlflow, restx
- **MSc in Signal/Information Processing & Learning:** (DIT, UoA) 2015-2018
2-year Master's degree, focused on Signal Processing and Machine Learning. Grade 8.78.
Specialized in Machine Learning, Signal and Multimedia Analysis and Processing, Ontology Engineering
> courses: {matlab, java, C++}, thesis: {python, tensorflow, keras, DCNN, LSTM}
- **BSc in Computer Science:** (DCSE-UoI) 2005-2013
4-year Bachelor's degree (Ptychion), with a focus on Machine Learning, AI and VLSI. Grade 7.08.
Specialized in Machine Learning and Artificial Intelligence, Multimedia Processing, Hardware Design
> courses: {C, C++, java, matlab, opengl, vhdl}, thesis: {matlab, UPSO}

EXPERIENCE



- **Postdoctoral Researcher:** [IIT-NCSR](#)D - Athens, Greece Nov 2021 - present
 - **R&D on Text ML:** Research and ML engineering on argumentation mining with transformer models.
> python, pytorch, huggingface, BERT, flair, elasticsearch, fastapi, docker
- **Machine Learning Engineer:** LANGaware June 2021-present
Machine Learning engineer, LANGaware company².
 - **ML Engineering:** Pipeline development / maintenance in the domain of multimodal neurodegenerative disease prediction and classification.
 - **ML Interpretability:** Work on feature / instance-based classification interpretation and presentation.
 - **ML Prototyping:** Work with the Data Science team to implement new solutions, experiments and prototypes.
 - **Technical Writing:** Authorship of proposals, reports and applications related to scientific/research projects.
> python, sklearn, RF, numpy, spacy, nltk, librosa, gcp
- **Machine Learning Engineer:** SciFY November 2020-August 2021
Machine Learning engineer, "Science For You" non-profit company³.
 - **ML Engineering:** Lead development and maintenance of text processing and classification pipelines in the legal domain. Evaluation, testing and optimization of developed solutions.
> python, sklearn, numpy, nltk, pytorch, transformers, BERT, flask
- **Research associate:** ILSP-ATHENA RC February-December 2019
Lead developer and researcher for the "Data for Impact" project, at the Institute of Language and Speech Processing, of the Athena Research Centre (ARC).

¹S. Niarchos Industrial scholarships, 2017

²<https://langaware.com/>

³<http://www.scify.gr/site/en/>

- **Text Machine Learning Research:** Lead developer on multimodal topic modelling on scientific publications. System maintenance and extension, administration and deployment.
 - **Result analysis and report:** Meta-analysis, data aggregation, visualization and presentation; indicator engineering and impact assessment. Project report compilation and authoring
 - **System / database administration:** VM / system management and setup, database extension and optimization.
 - java, mallet, LDA, spring, postgres, python, sklearn, django, flask
- **Research associate:** IIT-NCSR June 2016-September 2021
Lead developer and researcher at the “BigDataEurope” and “Storybot” projects, and during the PhD duration, focusing on ML methods over text, including but not limited to:
 - **R & D on Event Detection:** Lead developer on text & social media mining, processing and event detection systems. Maintenance, extension, scaling to big data, optimization and deployment. Research on event detection, document clustering and summarization, recommendation, entity resolution and personalization systems. Development and deployment of relevant systems.
 - **System administration & engineering:** VM and software virtualization containers design, deployment and maintenance. RESTful API frontend/backend development, website deployment and migration.
 - **Academic Publications:** Authorship and management of publications, large-scale experiment planning, execution, automation and analysis. Reviewer on machine learning scientific conference submissions .
 - **Community management:** Lecture group planning, administration, organization and mailing list management of the Content Analysis and Knowledge Technologies (CAKT) and Representation Analysis (REPR) research groups. Presentations on machine learning on school educational visits. Organizational and technical support for the MultiLing 2017 & 2019 conferences.
 - **Technical support:** Software development, support and technical assistance in multiple other institute projects.
 - java, jax-rs, pyinsect, NGG, cassandra, spark, bash, glassfish
- **Research associate:** ITI-CERTH February 2014 - September 2015
Lead developer and researcher at the “Linked TV: Television linked to the web” and “ForgetIT” projects, at the Information Technologies Institute (ITI), of the National Center for Research and Technology, Hellas (CERTH). Themi, Greece . Lead developer and researcher, focusing on video concept detection (VCD).
 - **Visual descriptors Research:** Research focusing on binary descriptors, colour extensions, normalization and clustering. Lead developer on a VCD pipeline end-to-end (feature extraction, processing, classification, visualization) modular design, implementation from scratch, maintenance and expansion and optimization. Spatiotemporal optimization.
 - **Deep Learning Research:** Research on deep Learning with deep convolutional neural networks (DCNNs), focusing on training, fine-tuning, network topology and architecture optimization, transfer and meta-learning. Lead developer of a DCNN-based real time VCD system, including design, development from scratch, maintenance, expansion and optimization.
 - **Academic Publications:** Authorship and management of publications, large-scale experiment planning, execution, automation and analysis. Reviewer on machine learning scientific conference submissions.
 - **Technical support:** Software development, support and technical assistance in multiple other institute projects.
 - C++, matlab, opencv, libsvm/linear, caffee, matconvnet, DCNN, openmp, cuda
- **Freelancer:** Naxos, Ioannina, Thessalonika, Athens 2000-2013
Freelance jobs, including tutoring on physics and mathematics, undertaking computer science university projects assignments, Greek to / from English translations, automobile dealership secretary, speedboat operator and others. Naxos, Athens, Ioannina, Thessalonika (2000-2013).

ACADEMIC PROJECTS

- **PhD projects:**
 - ◇ Beyond Deep Learning: Enriching Data Representations for Machine Learning Tasks. Supervised by P. Stamatopoulos, V. Karkaletsis, M. Koubarakis.

- Hate Speech Detection framework: Research findings during the PhD were realized into an industrial grade Hate Speech Detection solution, supporting training, testing and evaluation, hyper-parameter tuning, model deployment and tracking.
- **Selected MSc course projects:**
 - ◊ Thesis: Multimodal video classification using deep neural networks ([link](#)). Supervised by T. Giannakopoulos and S. Perantonis. Grade 10.0.
 - Semantic data and geo/spatiotemporal data handling, ontology engineering. (*Knowledge Technologies*)
 - Design, development, implementation and optimization of voice user interfaces. (*Voice Technologies*)
 - Implementation, tuning and large-scale experiments with multiple classification models (e.g. RBF, Bayesian and neural nets) in Weka and Matlab. Implementation of the RLS, Robbins-Monroe and LS denoising algorithms. Study and presentation of an AAAI-2016 conference submission. (*Pattern Recognition, Machine Learning, Advanced AI*)
 - Implementation of visual feature extractors, statistical analysers and classification algorithms. Design, implementation, documentation and IEEE-style paper authorship on the Stauffer-Grimson background subtraction method. (*Medical image analysis and image processing*)
 - Bibliographical research study on microcontroller and microcomputer platforms. (*Real time systems*)
 - Implementation of OO design patterns, multi-threading, aspect-oriented approaches and memory visualization. GC and memory management performance tests on Java and C++. Study and presentation of an OOPSLA'16 submission. (*Advanced programming techniques course*)
- **Selected BSc course projects:**
 - ◊ Thesis: Image registration using unified particle swarm optimization ([pdf](#)). supervisors: C. Nikou, K. Parsopoulos. Grade 10.0.
 - Implementation of classifiers (e.g. SVM, Bayes, K-means, MLPs). Implementation of and various heuristic and global optimization search algorithms. (*Pattern recognition, computational intelligence, AI courses*)
 - Implementation, design and documentation of graphical class analyser tools, a full compiler for a toy language, relational and ER-modeled databases. (*Software engineering, OO programming, compilers, database systems courses*)
 - Implementation of image processing tools, video codecs, video restoration and compression scripts, 2D and 3D games in OpenGL and X11, drawing GUI applications. (*Digital image processing, multimedia, computer graphics courses*)
 - Design, simulation and analysis of digital and analog circuits, up to a very large integration scale. Simulation of microprocessors and assembly programming. (*Digital design, microelectronics, integrated digital circuits, VLSI systems, computer architecture courses*)
- **Seminars and workshops:**
 - AGI Safety Fundamentals Course (Effective Altruism Cambridge, virtual, Jan-Mar 2022)
 - AI Safety Camp 2022 (virtual research workshop, Jan - Jun 2022)
 - “Stanford Existential Risk Initiative” Conference, 2022, virtual (Winter 2022)
 - Seminars, lectures and presentations, CAKT & SKEL research groups (IIT-NCSR, July 2016 - present).
 - “Object-centric machine learning”, seminar with L. Guibas from Stanford University (DIT-UoA, May 2016).
 - “A Value-Based Approach to Hardware Acceleration of Deep Learning”, seminar with A. Moshovos from Univ. of Toronto (DIT-UoA, May 2016).
 - “Image processing with MATLAB”, seminar with Mentor Hellas corp. (DIT-UoA, February 2016).
 - “CretaMASSS-2013 / HAISS'13-Agents” summer school, on multi-agent systems & AI (TUoC, Summer 2013).

- “Deep Learning in NLP” summer school (RANLP 2019, Fall 2019).
- “Open Source AI Workshop”, workshop on AI, Eclipse Foundation (virtual event, Summer 2020).
- “Artificial Intelligence for Multilanguage Services for Citizens and Businesses”, workshop, ELRC(virtual event, Winter 2020).
- “Prometheus Technology Transfer program”, workshop, SciFY, SRI, US Embassy(virtual event, April 2021).

PUBLICATIONS



• Journal publications:

1. N. Pittaras, G. Giannakopoulos, G. Papadakis, V. Karkaletsis, “Text classification with semantically enriched word embeddings”, JNLE Special Issue: Informing Neural Architectures for NLP with Linguistic and background Knowledge, 2020 ([pdf](#))
2. F. Markatopoulou, V. Mezaris, N. Pittaras, I. Patras, “Local Features and a Two-Layer Stacking Architecture for Semantic Concept Detection in Video”, IEEE TETC, 2015 ([ieee](#)) ([pdf](#)).

• Conference publications:

1. M. El-Haj, M. Litvak, N. Pittaras, G. Giannakopoulos, “The Financial Narrative Summarisation Shared Task (FNS 2020)”, COLING 2020, Winter 2020 ([pdf](#))
2. M. El-Haj, V. Athanasakou, S. Ferradans, C. Salzedo, A. Elhag, H. Bouamor, M. Litvak, P. Rayson, G. Giannakopoulos, N. Pittaras “Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation”, COLING 2020, Winter 2020 ([pdf](#))
3. M. El-Haj, A. AbuRa’ed, M. Litvak, N. Pittaras and G. Giannakopoulos, “The Financial Narrative Summarisation Shared Task (FNS 2020)”, FNP-FNS 2020, Barcelona, Spain (virtual) ([pdf](#)).
4. C. Vasilopoulou, N. Pittaras and G. Giannakopoulos, “A study of different, personalised representations of multi-omic data for bladder cancer tissue prediction”, SETN 2020, Athens, Greece (virtual).
5. A. Kosmopoulos, A. Liapis, G. Giannakopoulos and N. Pittaras, “Summarizing Game Reviews: First Contact”, SETN 2020, Athens, Greece (virtual) ([pdf](#)).
6. G. Papadakis, L. Tsekouras, M. Thanos, N. Pittaras, G. Simonini, D. Skoutas, P. Isaris, G. Giannakopoulos, T. Palpanas, M. Koubarakis, “JedAI3: beyond batch, blocking-based Entity Resolution”, EDBT 2020, Spring 2020 ([pdf](#)).
7. G. Giannakopoulos, N. Pittaras “The Summary Evaluation Task in the MultiLing-RANLP 2019 Workshop”, Multiling Workshop, RANLP2019, Varna, Bulgaria ([proceedings](#)).
8. N. Gialitsis, N. Pittaras, P. Stamatopoulos, “A topic-based sentence representation for extractive text summarization”, Multiling Workshop, RANLP2019, Varna, Bulgaria ([proceedings](#)).
9. N. Pittaras, V. Karkaletsis, “A study of semantic augmentation of word embeddings for extractive summarization”, Multiling Workshop, RANLP2019, Varna, Bulgaria ([proceedings](#)).
10. C. Themeli, G. Giannakopoulos, N. Pittaras “A study of text representations for Hate Speech Detection”, CICLING 2019, La Rochelle, France (to appear).
11. N. Pittaras, G. Papadakis, G. Stamoulis, G. Argyriou, E. K. Taniskidou, E. Thanos, G. Giannakopoulos, L. Tsekouras, E. Koubarakis, “GeoSensor: Semantifying Change and Event Detection over Big Data”, SAC 2019, Limassol, Cyprus ([pdf](#)).
12. N. Kostagiolas, N. Pittaras, C. Nikolaou, G. Giannakopoulos, “Exploring different sequence representations and classification methods for the prediction of nucleosome positioning”, bioRxiv, 2018 ([pdf](#)).
13. N. Pittaras, G. Giannakopoulos, L. Tsekouras, I. Varlamis, “Document clustering as a record linkage problem”, DocEng’18, Halifax, Nova Scotia, Canada, 2018 ([ResearchGate](#)).

14. G. Argyriou, G. Papadakis, G. Stamoulis, E. K. Taniskidou, N. Pittaras, G. Giannakopoulos, S. Albani, M. Lazzarini, E. Angiuli, A. Popescu, A. Argyridis and M. Koubarakis. “GeoSensor: On-line Scalable Change and Event Detection over Big Data”, The Web Conference 2018, Lyon, France, April 2018. ([pdf](#)).
15. S. Auer et al. “The BigDataEurope Platform - Supporting the Variety Dimension of Big Data”, International Conference on Web Engineering, Cham, 2017 ([springer](#)).
16. N.Pittaras, F.Markatopoulou, V.Mezaris, I.Patras, “Comparison of Fine-tuning and Extension Strategies for Deep Convolutional Neural Networks”, 23rd International Conference on Multimedia Modeling, Reykjavik, Iceland, Jan 2017 ([springer](#)) ([pdf](#)).
17. F. Markatopoulou, A. Ioannidou, C. Tzelepis, T. Mironidis, D. Galanopoulos, S. Arestis-Chartampilas, N. Pittaras, K. Avgerinakis, N. Gkalelis, A. Moutzidou, S. Vrochidis, V. Mezaris, I. Kompatsiaris, I. Patras, “ITI-CERTH participation to TRECVID 2015”, Proc. TRECVID 2015 Workshop, Gaithersburg, MD, USA, Nov. 2015 ([nist](#))([pdf](#)).
18. G. Kalpakis, T. Tsikrika, F. Markatopoulou, N. Pittaras, S. Vrochidis, V. Mezaris, I. Patras, I. Kompatsiaris, “Concept Detection on Multimedia Web Resources about Home Made Explosives”, Proc. Int. Workshop on Multimedia Forensics and Security (MFSec), held in conjunction with the 10th Int. Conf. on Availability, Reliability and Security (ARES), Toulouse, France, Aug. 2015 ([ieee](#)) ([pdf](#)).
19. F. Markatopoulou, N. Pittaras, O. Papadopoulou, V. Mezaris, I. Patras, “A Study on the Use of a Binary Local Descriptor and Color Extensions of Local Descriptors for Video Concept Detection”, Proc. 21st Int. Conf. on MultiMedia Modeling (MMM’15), Sydney, Australia, January 2015 ([springer](#)) ([pdf](#)).
20. N. Gkalelis, F. Markatopoulou, A. Moutzidou, D. Galanopoulos, K. Avgerinakis, N. Pittaras, S. Vrochidis, V. Mezaris, I. Kompatsiaris, I. Patras, “ITI-CERTH participation to TRECVID 2014”, Proc. TRECVID 2014 Workshop, Orlando, FL, USA, November 2014 ([nist](#)) ([pdf](#)).

- **Book chapters:**

1. N. Pittaras, S. Montanelli, G. Giannakopoulos, A. Ferrara, V. Karkaletsis, “Crowdsourcing in Single-document Summary Evaluation: the Argo Way”, Multilingual Text Analysis: Challenges, Models, and Approaches, World Scientific Publishers, 2020 ([amazon](#)).
2. G. Giannakopoulos, G. Kiomourtzis, N. Pittaras, V. Karkaletsis, “Scaling and semantically enriching language-agnostic summarization”, Trends and Applications of Text Summarization Techniques, 2019 ([IGI Global](#)).

ACADEMIC ACTIVITIES



- Course Assistant
 - Data Structures and Programming Techniques, DIT, UoA, Spring semesters 2016, 2017).
 - Object-Oriented Programming, DIT, UoA, Spring semester 2018).
 - Software development for Networks, DIT, UoA, Winter semester 2018, 2019).
 - Data Mining Techniques, DIT, UoA, Spring semesters 2019, 2020).
- Co-supervisor
 - BSc / MSc theses / Internships related to Machine Learning provided by G. Giannakopoulos / NCSRD
 - Theoretical / technical support and guidance, group coordination and organization
 - Scientific lectures and presentations on the CAKT / REPR research groups
- Reviewer on Machine learning, Natural language processing and Computer vision conferences / journals:
 - ICTAI, MultiLing, SLSP, CSL, ACL, ACL-RR, Comput. Intell., SETN, AACL-IJCNLP, Machine Learning, Electronics
- Conferences / Events

- MultiLing 2019: Workshop organizer, coordinator and administrator (RANLP 2019, Varna, Bulgaria).
- AI4EU: World Caffe rapporteur (philosophy and ethics table), IT support (Athens, Greece).
- FNP2020 and FNP-FNS2020: Publicity chair, task organizer (COLING 2020, virtual).
- FNP2021 and FNP-FNS2021: Publicity chair, task organizer (Lancaster University, UK).
- FNP2022 and FNP-FNS2022: Publicity chair, task organizer (LREC 2022).
- ACL-IJCNLP 2020: Program Committee (Suzhou, China).
- SLT Workshop 2021: Program Committee (virtual).
- SDPRA 2021: Program Committee (virtual).

TECH SKILLS



- Proficient with:
 - *programming*: C, C++, Java, Matlab, python, bash, L^AT_EX
 - *ML & DS*: pytorch, tensorflow, transformers, (num/sci)py, sklearn, ray-tune, mlflow, caffe, matconvnet, lib(svm/linear), opencv
 - *IDEs, frameworks & SDKs*: IDEs, NetBeans, Code::Blocks, qtCreator, Visual Studio 2010 / 2013, IDEA, CLion, PyCharm, sublime-text, VScode, maven
 - *big data & databases*: Cassandra, SPARK, MYSQL, POSTGRESQL
 - *semantic web*: RDF, (geo/st)SPARQL, OWL
 - *team and documentation*: Scrum, Kanban, Jira, Trello, Confluence, sphinx
 - *other*: git, docker, linux, ffmpeg
- Familiar with:
 - *programming*: scala, perl
 - *web* : HTML, CSS, javascript, apache (http server/tomcat), glassfish, liferay, hugo, flask, Jersey, restx
 - *APIs & libraries*: X, THREE.js, CUDA, Weka, OpenGL, glut, SDL2, android, django, sphinx
 - *IDEs & SDKs*: GCP, Eclipse, DevC++, kDevelop, Godot, protege, Android SDK, Unity 3D, Blender, CSLU RAD, IBM Websphere Voice Toolkit
 - *circuit design and simulation*: VHDL, Altera Quartus II, Capture CIS, Pspice
 - *databases & cloud*: mongoDB, GCP
 - *other*: QGIS, gdb

LANGUAGE SKILLS



- Greek (Mother tongue).
- English (Fluent: Cambridge FCE, CAE and CPE (grades A, A and C respectively)).
- German, French (Basic).

MISCELLANEOUS SKILLS



- Licensed to drive cars and speedboats.
- Completed lifeguard theoretical and practical training at PA.S.X.NA. lifeguard school.

HOBBIES AND INTERESTS



- Long distance running and cycling, swimming.
- Musical instruments (guitar, piano, some bass and ocarina), singing, song writing.
- Philosophy of science, AI safety, popular science (number theory, particle physics).
- Video games & indie game development.