

Jaime G. Carbonell -- Curriculum Vita

Business Address

Language Technologies Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213 USA
Telephone: (412) 268-7279
Email: jgc@cs.cmu.edu

Fax: (412) 268-6298
<http://www.cs.cmu.edu/~jgc/>

Citizenship: USA

Academic and Professional Positions Held

2012 - Present	University Professor, School of Computer Science, Carnegie Mellon University
1996 - Present	Director, Language Technologies Institute, Carnegie Mellon University
1995 - Present	Allen Newell Professor of Computer Science, Carnegie Mellon University
2006 - Present	Adjunct Faculty, Dept of Computational Biology, U Pittsburgh Medical School.
2001 - Present	Co-Founder, Board chairman, Carnegie Speech Corporation
1995 - 2003	Co-Founder, Board chairman, Wisdom Technologies Corporation
1986 - 1996	Director, Center for Machine Translation, Carnegie Mellon University
1987 - 1995	Professor of Computer Science, Carnegie Mellon University
1983 - 1987	Associate Professor of Computer Science, Carnegie Mellon University
1983 - 1996	Co-Founder, Director, Scientific advisor, Carnegie Group Inc.
1979 - 1983	Assistant Professor of Computer Science, Carnegie Mellon University
1975 - 1978	Research Fellow in Computer Science, Yale University
1975 - 1977	Teaching Assistant in Computer Science, Yale University
1974 - 1975	Research Assistant, Center for Space Research, MIT
1971 - 1975	Research Programmer, Artificial Intelligence, BBN, Cambridge, MA

Education

1975 - 1979	Yale University	PhD 1979 - Computer Science MPhil 1977 - Computer Science MS 1976 - Computer Science
1971 - 1975	Massachusetts Institute of Technology	SB 1975 - Physics SB 1975 - Mathematics

Professional Organizations and Activities

Member: ACM (elected Chair of SIGART 1983-85), ACL, AAAI (AAAI Fellow 1988-present, AAAI executive committee 1990-92), Cognitive Science Society, Sigma Xi, AMTA.

Government and other Committees: NSF/CISE Scientific Advisory Committee (2010-2014), NIH Human Genome Scientific Advisory Committee, aka "Watson Committee" (1988-1992), DFKI (National German AI Lab) Scientific Advisory Board (1988-2003), National Institute of Standards and Technology (NIST/IAD), Scientific Advisory Committee (1997-2001), DOE Oakridge National Laboratories Scientific Advisory Committee (1985-1987), GMD/PSI Information Sciences (Germany) Scientific Advisory board (1990-2001), Citigroup Technology Advisory committee (1987-1995).

Book Series co-Editor: *Lecture Notes in Artificial Intelligence* North American editor, Springer (1996-2008).

Editorial Boards: Machine Learning Journal (1984-2000, editor-in-chief 1988-1993), Machine Translation Journal (1980's), Artificial Intelligence Journal (1984-2008).

Other Accomplishments

Research: Invented multiple well-known algorithms and methods in statistical machine learning including: Proactive Machine Learning (with Donmez) for multi-source cost-sensitive active learning, Linked Conditional Random Fields (L-SCRF, with Liu) for predicting tertiary and quaternary protein folds, new regularization-based transfer learning methods (with Kshirsagar), Maximal Marginal Relevance (MMR, with Goldstein) for information novelty, retrieval and summarization, topic-conditioned modeling for novelty detection, symmetric optimal phrasal

IPR2018-01079, Exhibit 2006

alignment method for trainable example-based and statistical machine translation, series-anomaly modeling for financial fraud detection and syndromic surveillance, knowledge-based interlingual machine translation (and later extended it with colleagues: Tomita, Nirenburg, Nyberg, and Mitamura), transformational analogy for case-based reasoning, derivational analogy for reconstructive justification-based reasoning (with Veloso), robust case-frame parsing (with Hayes), Seeded Version-Space Learning (with polynomial complexity), and developed improvements to several other machine learning algorithms. Current research foci include robust statistical learning and mapping protein sequences to 3D structure and inferring functional properties, automated transfer-rule learning for Machine Translation, , enriched active transfer learning context-based machine translation, and machine translation for very rare languages.

Software artifacts: Protein structure prediction software (with Liu), active and proactive learning algorithms (with Donmez), PRODIGY general-purpose planner (with Minton and Veloso), KANT knowledge-based machine translation (with Nyberg and Mitamura), JAVELIN Question-Answering system (with Nyberg et al), CONDOR search engine (with Callan et al), MMR summarizer engine (with Goldstein), POLITICS simulated reasoning engine, DIPLOMAT speech-MT (with Frederking et al), EBMT example-based general purpose MT (with Brown), SMOKEY distributed sensor-based fire detection and suppression system, and several others.

Education: Created the PhD program at CMU on Language Technologies with 100's of PhD graduates. Designed new courses in Language Technologies, Machine learning, Data Mining and eCommerce. Edited 3 books on Machine Learning. Taught undergraduate and graduate courses in Artificial Intelligence, Machine Learning, Software Engineering, Natural Language Processing, Language Technologies, Data Mining, Information Retrieval, Web-Based Architectures, Machine Translation, Algorithms, and related Computer Science topics.

Other initiatives: Created the Center for Machine Translation at CMU in 1986, and the Language Technologies Institute in 1996. Co-creator & co-director of the Universal Library and million-book project (US, China, India). MT Summit chair 1991, Co-designer of interactive pinpointing speech tutor. Co-creator & co-PI of new Computational Bilingualism Initiative at CMU. Helped create the CMU-Pitt Computational Biology Program.

Industrial Consulting: Meaningful Machines (participated in designing the context-based machine translation methods, 2002-2009), Industrial Scientific (data mining to improve workplace safety, 2008-2017), Carnegie Speech Inc. (business and technology strategy for intelligent language tutoring, 2002-present) Carnegie Group Inc (expert systems & financial data mining: 1984-1997), Citicorp (financial data mining, real time transactions, new IT technology insertion: 1990-1998), Lycos Inc. (launching the internet search engine 1998-1999), Vivisimo (2003-2006, scientific advisory board), Searchline (launched the new search engine, 2002-2004), Wisdom Technologies (financial optimization and data mining for corporate treasuries, large banks, etc. 1995-2003), Dynamix Technologies (designed their large-scale data mining engine with applications to Homeland Security, AR/AP financial optimization, transactional optimization in bond trading and in re-insurance, pattern discovery in network-wire-transfer streams, syndromic surveillance, etc. 2000-present), Peak Strategy (financial analytics, investment, mining, 2005-2007), Boeing (designed their intensive data mining course), Citibank (technology advisory board for 10 years, focusing on text and data mining, fraud detection, optimization methods, etc.), plus about 10 other shorter-term engagements in data mining in industry, including financial data mining: transactional fraud detection, instant-credit bad-debt minimization, collection effectiveness, cash-flow optimization (sweep/invest/float), data mining for risk/cost tradeoff minimization (FX, interest rate, counterparty, debt, etc), and so on. During these engagements process I designed or co-designed three data mining engines, and evaluated and help improve numerous other data mining engines, including the machine learning algorithms at their core. Consultant for UPMC, Precyse and Nthrive leading ICD9, ICD10 automated coding of EMRs vis NLP and ML.

Miscellaneous: Gave over 500 invited or refereed-paper presentations (colloquia, seminars, panels, addresses, conferences, key-notes, etc.). Received the *IJCAI Best Paper Award* in 1997 for translingual information retrieval. Received the CMU Computer Science Department's teaching award (1987), the Sperry Fellowship for excellence in AI research (1986), unsolicited gifts from Alcoa and Hughes corporations for research in machine learning, and a "recognition of service" award from the ACM for the SIGART presidency (1983-1985). Provided congressional testimony on machine translation (1990). Leader of the NSF/JTEC study group on Japanese machine translation science and technology (1990-91). Principal investigator or major participant in research grants and contracts totaling over 100 million dollars since 1979 at CMU..

Expert witness: Provided expert testimony and depositions, expert reports, advice, consulting, etc. in several cases involving intellectual property including: software design and copyrights, patents for data mining, patents for search engine algorithms and methods, speech recognition, integrated software systems, computational and statistical applications in finance, statistical machine learning, text mining, text processing, natural language, etc.

Patents: Inventor in several issued patents, provided advice on patentability of new software and methods, helped draft software-related patents, conducted prior art searches, composed responses to patent examiners, evaluated validity of issued patents. As expert witness provided opinions and reports on patent infringement, patent validity (anticipation, obviousness), claim constructions, Inter Partes Reviews, etc.

Completed PhD Advisees & co-Advisees (and first job after graduation)

1. Masaru Tomita, Professor and Dean, Keio University, Computer Science and Biology
2. Steven Minton, USC-ISI & CTO of Fetch Inc
3. Manuela Veloso, Professor, CMU Computer Science
4. Jack Mostow, Research Professor, CMU Robotics
5. Michael Mauldin, Founder of Lycos Inc. (and former faculty at CMU)
6. Oren Etzioni, Professor, U of Washington (and founder of Metacrawler)
7. Robert Frederking, Sr. System Scientist, CMU Language Technologies Institute
8. Alex Hauptmann, Research Professor, CMU Computer Science
9. Klaus Gross, Chief Scientist, Haley Enterprises
10. Wei-Min Shen, USC-ISI and faculty at USC Computer Science
11. Eric Nyberg, Professor, CMU Language Technologies Inst
12. Ralf Brown, Sr. System Scientist, CMU Language Technologies Institute
13. Craig Knoblock, USC-ISI and faculty at USC Computer Science
14. Akira Ushioda, Fujitsu Labs, Manager Language Technologies
15. Mark Perlin, CEO Cybergenetics
16. Daniel Kuokka, Senior researcher, Tektronix
17. Yolanda Gil, USC-ISI, senior researcher
18. Xuemei Wang, Lockheed Martin & ISLE Computing
19. Robert Joseph, Verizon Communication, senior researcher
20. Jill Fain-Lehman, Founder of Autism Technology, Disney Pittsburgh
21. Eugene Fink, Senior Sys Scientist CMU (formerly Asst Prof, U of Florida)
22. Angela Kennedy, President Carnegie Speech (formerly CEO Wisdom Technologies)
23. Alicia Perez, Faculty at Universidad Catolica de Salta, Argentina
24. Jim Blythe, Researcher, USC-ISI
25. Alex Franz, Google Inc.
26. Yan Qu, Senior Researcher, Oracle Corp
27. Jade Goldstein, US DoD research laboratory
28. Boyan Onyskevich, US DoD research laboratory, senior manager
29. Michael Mateas, Faculty, University of CA.
30. Kathreina Probst, Faculty at Georgia Tech
31. Yan Liu, Faculty at University of Southern Cal.
32. Lucian Lita, Siemens Princeton Research Lab
33. Paul Bennett, Microsoft Research Labs
34. Chun Jin, AT&T Research
35. Ariadna Font-Llujos, IBM, Big Data Labs
36. Ulas Bardak, Postdoctoral fellow, Tokyo University
37. Christian Monson, Oregon Graduate Institute
38. Cenk Gazen, Fetch Inc. Los Angeles
39. Monica Rogati, Linked-In, manager of R&D
40. Vasco Pedro, CEO of Bueda Inc.
41. Meryem Pinar Donmez, Salesforce R&D
42. Jingrui He, Faculty, Stevens Institute
43. Jae-Dong Kim, Google Pittsburgh
44. Rashmi Gangadharaiah, Microsoft Research Labs
45. Oznur Tastan, Microsoft Research Labs, then Bilkent U. Faculty.
46. Jonathan Elsas, Google Pittsburgh
47. Vamshi Ambati, AT&T Research
48. Bin Fu, Google, NYC
49. Mehrbod Sharifi, Google Pittsburgh
50. Ravi Starzl, CMU Faculty
51. Xi Chen, UC Berkeley Postdoc, then NYU Faculty
52. Balakrishnan, Sevaraman, UC Berkeley Postdoc, then CMU Faculty
53. Liu Yang, CMU Postdoc, then Princeton Postdoc
54. Mohit Kumar, Flipkart Inc.
55. Selen Uguroglu, Apple Inc.
56. Luis Marujo, Snapchat.
57. Meghana Kshirsagar, IBM Research
58. Andrew Hsi, Bloomberg
59. Ashiqur, Kurbadash, CMU Postdoc.
60. Shane Moon, Facebook

61. Keerthiriam Murugesan, IBM Research
62. Jesse Dunietz, MIT
63. Guoqing Zheng, Google
64. Jeff Flanigan, UMass

Current PhD advisees:

1. Daegun Won, CMU, LTI
2. Wan Li, CMU, LTI
3. George Phillipp, CMU, Comp Sci Dept
4. Adams Wei-Yu, CMU, Machine Learning Dept.
5. Petar Stojano, CMU, Computer Sci Dept
6. Jay-Yoon Lee, CMU, Computer Sci Dept
7. Zirui (Ed) Wang, CMU, LTI
8. Jiateng Xie, CMU, LTI
9. Shruti Rijhwani, CMU LTI
10. Sanket Mehta, CMU LTI
11. Hieu Pham, CMU LTI
12. Aditya Chandrasekar, CMU LTI
13. Vidhisha Balachandran, CMU LTI

Publications

1. Lee, J., Sanket, Mehta, S., Wick, M., Tristan, J. and Carbonell, J., "Gradient-based Inference for Networks with Output Constraints", *Proc of AAAI*, Hawaii, USA, 2019.
2. Stojanov, P., Zhang, K., Gong, M., Carbonell, J., "Low-Dimensional Density Ratio Estimation for Covariate Shift Correction", *Proc. of AISTATS*, 2019.
3. Stojanov, P., Zhang, K., Gong, M., Carbonell, J., "Data-Driven Approach to Multiple-Source Domain Adaptation", *Proc. of AISTATS*, 2019.
4. Rijhwani, S., Jiateng Xie, J., Neubig, G., Carbonell, J., "Zero-shot Neural Transfer for Cross-lingual Entity Linking", *Proc of AAAI*, Hawaii, USA, 2019.
5. KhudaBukhsh, A., Carbonell, J., "Expertise Drift in Referral Networks", In *Proc of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2018.
6. KhudaBukhsh, A., Carbonell, J., and Jansen, J., "Robust Learning in Referral Networks: A Comparative Analysis", *Journal of Intelligent Information Systems*, June, 2018.
7. Yang, L, Hanneke, S., and Carbonell, J. "Bounds on the Minimax Rate for Estimating a Prior over a VC Class from Independent Learning Tasks.", *Theoretical Computer Science*, Vol. 716, pp. 124-140, 2018 (Extended version of ALT 2015 paper.)
8. Wang, Z. and Carbonell, J., "Towards More Reliable Transfer Learning", *Proc. of ECML/PKDD*, 2018.
9. Xie, J., Z Yang, Z., Neubig, G., Smith, N., Carbonell, J., "Neural cross-lingual named entity recognition with minimal resources", *Proc. of EMNLP*, 2018.
10. Chaudhary, A., Zhou, C., Levin, L., Neubig, G., Mortensen, D., and Carbonell, J., "Adapting Word Embeddings to New Languages with Morphological and Phonological Subword Representations", *Proc. of EMNLP*, 2018.
11. Metha, S., Lee, J. and Carbonell, J., "Towards Semi-Supervised Learning for Deep Semantic Role Labeling" In *Proc of Empirical Methods for Natural Language Processing (EMNLP)*, 2018.
12. Dunietz, J., Levin, L., and Carbonell, J. "DeepCx: A Transition-Based Approach for Shallow Semantic Parsing with Complex Constructional Triggers", In *Proc of Empirical Methods for Natural Language Processing (EMNLP)*, 2018.
13. Won, D., Jansen, P., and Carbonell, J., "Temporal Transfer Learning for Drift Adaptation", In *Proc. of the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 2018.
14. Philipp, G., Song, D., and Carbonell, J. "Gradients explode – Deep Networks and Shallow-ResNet Explained", ICLR workshop, 2018.
15. Philipp, G., and Carbonell, J., "The Nonlinearity Coefficient-Predicting Overfitting in Deep Neural Networks", arXiv preprint arXiv:1806.00179, 2018.
16. Wang, Z. and Carbonell, J. "Towards more Reliable Transfer Learning", Preprint in arXiv:1807.02235, 2018.
17. Bourai, A., and Carbonell, J., "I Know What You Don't Know: Proactive Learning through Targeted Human Interaction." In *Proc of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2018.

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.