Jon Paul Janet

Research CV

I am a scientist applying state-of-the-art machine learning techniques to solve challenging problems in early stage drug discovery. I am interested in how machine learning and computational chemistry can be combined to inform molecular design. My research has been featured in NeurIPS, Nature Machine Intelligence and various high-impact chemistry journals.

Experience

2022– **Associate Principal Scientist, Molecular AI**, AstraZeneca.
develop[1, 2] and apply generative reinforcement learning[3] in combination with traditional structure[4, 5]- and ligand-based[6] computational chemistry techniques to commercial early drug discovery projects. I also work with molecular property prediction[7–9], transformers for reaction informatics[10], active learning to close the gap between simulation and generative modeling, and generally evangelize around data-driven drug discovery!

2020–2022 **Senior Scientist, Computational Chemistry & Machine Learning**, AstraZeneca.

2015–2019 **Research Assistant, Machine Learning Materials Design**, Massachusetts Institute of Technology.
applied techniques from applied mathematics and machine learning combined with quantum chemical simulation for rational exploration and optimization in chemical space [11–27]. Featured in MIT News a, ACS Chemical & Engineering News b, c and MIT Energy Futures d and authored a book [28] titled Machine Learning in Chemistry as part of the ACS In Focus Series.

2018 **Teaching Assistant**, *Massachusetts Institute of Technology*.
graduate computational chemistry course at MIT, 6.8/7 student rating

2017 **Teaching Assistant**, *Singapore University of Technology and Design*.
second-year differential equations and optimization course for engineers

2014 **Internship, Computational Fluid Dynamics**, *Helmholtz-Zentrum Dresden-Rossendorf*.
investigated nucleation processes in dynamic 3D simulation of multiphase fluid flow[29]

2013 **Research and Teaching Assistant**, *University of the Witwatersrand*.
designed and conducted experiments on biological remediation of acid mine drainage[30]

2013 **Research Assistant**, *Process Modeling Group, University of Cape Town*.
designed reactor models for oil and gas processes

Spotlight Publications

1. D. Buterez, J. P. Janet, S. J. Kiddle, D. Oglic, and P. Liò. Graph neural networks with adaptive readouts. In A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, editors, *Advances in Neural Information Processing Systems*, 2022. doi: 10.48550/arXiv.2211.04952 – link
2. J. Guo, V. Fialková, J. D. Arango, C. Margreitter, J. P. Janet, K. Papadopoulos, O. Engkvist, and A. Patronov. Improving de novo molecular design with curriculum learning. *Nature Machine Intelligence*, 4(6):555–563, 2022. doi: 10.1038/s42256-022-00494-4 – link
3. J. P. Janet, C. Duan, A. Nandy, F. Liu, and H. J. Kulik. Navigating transition-metal chemical space: Artificial intelligence for first-principles design. *Accounts of Chemical Research*, 54(3): 532–545, 2021. doi: 10.1021/acs.accounts.0c00686 – link
4. J. P. Janet, S. Ramesh, C. Duan, and H. J. Kulik. Accurate multiobjective design in a space of millions of transition metal complexes with neural-network-driven efficient global optimization. *ACS Central Science*, 6(4):513–524, 2020. doi: 10.1021/acscentsci.0c00026 – link
5. J. P. Janet, C. Duan, T. Yang, A. Nandy, and H. J. Kulik. A quantitative uncertainty metric controls error in neural network-driven chemical discovery. *Chemical Science*, 10:7913–7922,

2019. doi: 10.1039/C9SC02298H – link

Education

2015–2019 **Ph.D. Chemical and Computational Engineering**, *Massachusetts Institute of Technology*.
Coursework in in *statistical learning, atomistic simulation, inference and information.* Thesis on *data-driven materials design* with Heather Kulik and Youssef Marzouk.

2014–2015 **M.Sc. Scientific Computing**, *Technical University of Berlin*, 1.0 "Sehr Gut".
Coursework in *control theory, differential algebraic equations (DAEs), optimal control, model order reduction.* Thesis in DAE controllability with Volker Mehrmann.

2013–2014 **M.Sc. Applied Mathematics**, *Royal Institute of Technology, Stockholm*, A grade.
Coursework in *numerical analysis, fluid mechanics, finite element methods, parallel programming*

2009–2012 **B.Sc. Chemical Engineering (Hons.)**, *University of Cape Town*, 3.95 GPA, 1st in class.
Core chemical engineering curriculum in *process design, transport, mathematics and modeling*

Skills and Projects

- **Programming**: High proficiency: Python, R, Matlab. Basic proficiency: C, MPI/high performance/parallel programming.
- **Machine learning**: Experience designing and training neural networks and kernel models and uncertainty quantification. Experienced with Pytorch, TensorFlow, Keras and H2O frameworks.
- **Computational Chemistry and simulation**: Experienced with quantum and classical simulation of molecular systems, dynamics, calculation automation and integration with machine learning, docking, free energy perturbation simulations.
- **General**: Expert proficiency with LaTeX, TikZ, MS Office. Experienced with Linux, Windows & supercomputer administration, hardware and maintenance.
- **Open-source projects**:
  1. Core developer of molSimplify for 5 years – a python toolbox for inorganic molecular assembly and prediction. In particular, implemented the first neural-network assisted geometry initialization for quantum chemistry.
  2. Founder, core developer of molSimplify Automatic design – a python toolbox for combining simulation and machine learning for molecular design

Awards and Fellowships

- *Awards*: **ACS Chemical Computing Group Excellence Award for Graduate Students 2019**, **MIT Center for Computational Engineering MathWorks Prize for doctoral research 2019**, South African Institution of Chemical Engineers Silver Medal 2012. City of Cape Town Engineering Medal 2012. Malan Prize Overall Medal 2012. Chemical Engineering Class Medal 2012, 2011. Organic and Inorganic Chemistry Medal 2010. Dean's Merit List, 2009–2012.
- *Fellowships*: MIT-SUTD Graduate Fellow 2017, Haas Family Fellowship 2015–2016. Erasmus Mundus Scholarship 2013–2015.

Talks and Workshops

All talk slides available at `https://jpjanet.io/`

- (invited) ACS In Focus invited webinar, 12/02/21. Machine Learning in Chemistry: Now and in the Future

- (invited) ACS In Focus invited webinar, 05/25/21. Machine Learning in Chemistry: Now and in the Future
- (contributed) AIChE Annual Meeting, Orlando 11/11/19. Multi-Objective, Machine-Learning Assisted First-Principles Design of Transition Metal Complexes for Redox Couples
- (invited) NSF TRIPODS+X Foundational & Applied Data Science for Molecular and Material Science & Engineering, 05/22/19. Hybrid machine-learning and first-principles design for transition metal complexes
- (contributed) American Chemical Society National Meeting, Orlando 03/31/19. Controlling generalization errors when using ML informed property models for discovery of molecular inorganic systems
- (invited) Presented a workshop on machine learning in chemistry, an introduction to ML for graduate students in chemical sciences at Ben Gurion University (02/14/19), Hebrew University of Jerusalem (02/17/19) and MIT (03/29/19)
- (contributed) American Chemical Society National Meeting, New Orleans 03/19/18. Machine learning for inorganic molecular design: descriptors and similarity in transition metal chemical space
- (contributed) AIChE Annual Meeting, Minneapolis 09/29/17. Mapping transition metal chemical space for machine learning models
- (contributed) American Chemical Society National Meeting, Washington 08/23/17. Mapping transition metal chemical space with continuous descriptors – feature selection and implications for machine learning models
- (contributed) New England Research Forum, Worcester 07/28/17. Describing transition metal chemical space for predictive machine learning

Contributed Posters and Exhibits

- Israel Chemical Society National Meeting, Tel Aviv 02/13/19. First principles inorganic materials design using machine-learned surrogate models.
- MIT Museum Energy Night, Boston 10/19/18. Interactive demonstration: machine learning for materials design
- American Chemical Society National Meeting, Boston 08/21/18. Accelerating inorganic discovery with machine learning and automation.
- MIT Energy Initiative Seed Fund Poster Session, Boston 02/26/18. Next Generation Quantitative Structure Property Relationships for Lubricants From Machine Learning and Advanced Simulation
- American Conference on Theoretical Chemistry, Boston 07/07/17. Training neural networks for transition metal complex screening and design

Academic Publications

[1] J. Guo, V. Fialková, J. D. Arango, C. Margreitter, J. P. Janet, K. Papadopoulos, O. Engkvist, and A. Patronov. Improving de novo molecular design with curriculum learning. *Nature Machine Intelligence*, 4(6):555–563, 2022. doi: 10.1038/s42256-022-00494-4.

[2] J. Guo, F. Knuth, C. Margreitter, J. P. Janet, K. Papadopoulos, O. Engkvist, and A. Patronov. Link-invent: Generative linker design with reinforcement learning. *Digital Discovery*, 2023. doi: 10.1039/D2DD00115B.

[3] J. P. Janet, L. Mervin, and O. Engkvist. Artificial intelligence in molecular de novo design: Integration with experiment. *Current Opinion in Structural Biology*, 80:102575, 2023. doi: 10.1016/j.sbi.2023.102575.

[4] J. Guo, J. P. Janet, M. R. Bauer, E. Nittinger, K. A. Giblin, K. Papadopoulos, A. Voronov, A. Patronov, O. Engkvist, and C. Margreitter. Dockstream: a docking wrapper to enhance de novo molecular design. *Journal of Cheminformatics*, 13(1):89, 2021. doi: 10.1186/s13321-021-00563-7.

[5] H. Moore, C. Margreitter, J. P. Janet, O. Engkvist, B. de Groot, and V. Gapsys. Automated relative binding free energy calculations: from SMILES to ΔΔG. *ChemRxiv*, 2022. doi: 10.26434/chemrxiv-2022-vqbxg.

[6] K. Papadopoulos, K. A. Giblin, J. P. Janet, A. Patronov, and O. Engkvist. De novo design with deep generative models based on 3d similarity scoring. *Bioorganic & Medicinal Chemistry*, 44:116308, 2021. doi: https://doi.org/10.1016/j.bmc.2021.116308.

[7] D. Buterez, J. P. Janet, S. Kiddle, and P. Liò. Multi-fidelity machine learning models for improved high-throughput screening predictions. *ChemRxiv*, 2022. doi: 10.26434/chemrxiv-2022-dsbm5.

[8] D. Buterez, J. P. Janet, S. Kiddle, and P. Liò. Mf-pcba: Multi-fidelity high-throughput screening benchmarks for drug discovery and machine learning. *ChemRxiv*, 2022. doi: 10.26434/chemrxiv-2022-cb3tz.

[9] D. Buterez, J. P. Janet, S. J. Kiddle, D. Oglic, and P. Liò. Graph neural networks with adaptive readouts. In A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, editors, *Advances in Neural Information Processing Systems*, 2022. doi: 10.48550/arXiv.2211.04952.

[10] J. P. Janet, A. Tomberg, and J. Boström. Reusability report: Learning the language of synthetic methods used in medicinal chemistry. *Nature Machine Intelligence*, 3(7):572–575, July 2021. ISSN 2522-5839. doi: 10.1038/s42256-021-00367-2.

[11] N. Arunachalam, S. Gugler, M. G. Taylor, C. Duan, A. Nandy, J. P. Janet, R. Meyer, J. Oldenstaedt, D. B. K. Chu, and H. J. Kulik. Ligand additivity relationships enable efficient exploration of transition metal chemical space. *The Journal of Chemical Physics*, 157(18):184112, 2022. doi: 10.1063/5.0125700.

[12] J. P. Janet. *Data-Driven Mapping of Inorganic Chemical Space for the Design of Transition Metal Complexes and Metal-Organic Frameworks*, chapter 7, pages 127–179. 2022. doi: 10.1021/bk-2022-1416.ch007.

[13] D. R. Harper, A. Nandy, N. Arunachalam, C. Duan, J. P. Janet, and H. J. Kulik. Representations and strategies for transferable machine learning improve model performance in chemical discovery. *The Journal of Chemical Physics*, 156(7):074101, 2022. doi: 10.1063/5.0082964.

[14] J. P. Janet, C. Duan, A. Nandy, F. Liu, and H. J. Kulik. Navigating transition-metal chemical space: Artificial intelligence for first-principles design. *Accounts of Chemical Research*, 54(3):532–545, 2021. doi: 10.1021/acs.accounts.0c00686.

[15] S. M. Moosavi, A. Nandy, K. M. Jablonka, D. Ongari, J. P. Janet, P. G. Boyd, Y. Lee, B. Smit, and H. J. Kulik. Understanding the diversity of the metal-organic framework ecosystem. *Nature Communications*, 11(1):4068, Aug 2020. doi: 10.1038/s41467-020-17755-8.

[16] J. P. Janet, S. Ramesh, C. Duan, and H. J. Kulik. Accurate multiobjective design in a space of millions of transition metal complexes with neural-network-driven efficient global optimization. *ACS Central Science*, 6(4):513–524, 2020. doi: 10.1021/acscentsci.0c00026.

[17] J. P. Janet, C. Duan, T. Yang, A. Nandy, and H. J. Kulik. A quantitative uncertainty metric controls error in neural network-driven chemical discovery. *Chemical Science*, 10:7913–7922, 2019. doi: 10.1039/C9SC02298H.

[18] S. Gugler, J. P. Janet, and H. Kulik. Enumerating de novo small inorganic complexes for machine learning and chemical discovery. *Molecular Systems Design & Engineering*, 5:139–152, 2020. doi: 10.1039/C9ME00069K.

[19] J. P. Janet, F. Liu, A. Nandy, C. Duan, T. Yang, S. Lin, and H. J. Kulik. Designing in the face of uncertainty: Exploiting electronic structure and machine learning models for discovery in inorganic chemistry. *Inorganic Chemistry*, 58(16):10592–10606, 2019. doi: 10.1021/acs.inorgchem.9b00109.

[20] C. Duan, J. P. Janet, F. Liu, A. Nandy, and H. J. Kulik. Learning from failure: Predicting electronic structure calculation outcomes with machine learning models. *Journal of Chemical Theory and Computation*, 15(4):2331–2345, 2019. doi: 10.1021/acs.jctc.9b00057.

[21] J. P. Janet, L. Chan, and H. J. Kulik. Accelerating chemical discovery with machine learning: Simulated evolution of spin crossover complexes with an artificial neural network. *The Journal of Physical Chemistry Letters*, 9(5):1064–1071, 2018. doi: 10.1021/acs.jpclett.8b00170.

[22] A. Nandy, C. Duan, J. P. Janet, S. Gugler, and H. J. Kulik. Strategies and Software for Machine Learning Accelerated Discovery in Transition Metal Chemistry. *Industrial & Engineering Chemistry Research*, 57(42):13973–13986, 2018. doi: 10.1021/acs.iecr.8b04015.

[23] J. P. Janet and H. J. Kulik. Predicting electronic structure properties of transition metal complexes with neural networks. *Chemical Science*, 8:5137–5152, 2017. doi: 10.1039/C7SC01247K.

[24] J. P. Janet and H. J. Kulik. Resolving transition metal chemical space: Feature selection for machine learning and structure–property relationships. *The Journal of Physical Chemistry A*, 121 (46):8939–8954, 2017. doi: 10.1021/acs.jpca.7b08750.

[25] J. P. Janet, T. Z. H. Gani, A. H. Steeves, E. I. Ioannidis, and H. J. Kulik. Leveraging cheminformatics strategies for inorganic discovery: Application to redox potential design. *Industrial & Engineering Chemistry Research*, 56(17):4898–4910, 2017. doi: 10.1021/acs.iecr.7b00808.

[26] A. Bajaj, J. P. Janet, and H. J. Kulik. Communication: Recovering the flat-plane condition in electronic structure theory at semi-local dft cost. *The Journal of Chemical Physics*, 147(19):191101,

2017. doi: 10.1063/1.5008981.
[27] J. P. Janet, Q. Zhao, E. I. Ioannidis, and H. J. Kulik. Density functional theory for modelling large molecular adsorbate–surface interactions: a mini-review and worked example. *Molecular Simulation*, 43(5-6):327–345, 2017. doi: 10.1080/08927022.2016.1258465.
[28] J. P. Janet and H. J. Kulik. *Machine Learning in Chemistry*. American Chemical Society, Washington, DC, USA, 2020. doi: 10.1021/acs.infocus.7e4001.
[29] J. P. Janet, Y. Liao, and D. Lucas. Heterogeneous nucleation in cfd simulation of flashing flows in converging–diverging nozzles. *International Journal of Multiphase Flow*, 74:106–117, 2015. doi: 10.1016/j.ijmultiphaseflow.2015.04.005.
[30] J. P. Janet, K. Harding, C. Sheridan, and D. Drake. Increasing pumping depth in the long-term management of acid mine drainage. In *WISA 2014: Water Institute of Southern Africa*, South Africa, 2014.