

Youhan Lee

Daejeon, S. KOREA | +82-10-9397-4570

bshz15@gmail.com | <https://www.linkedin.com/in/youhanlee/> |
<https://scholar.google.co.kr/citations?user=EFNg9UcAAAAJ&hl=en> |
<https://www.kaggle.com/youhanlee>

Experienced AI/ML research scientist & engineer with PhD in chemical and biomolecular engineering. Experienced material discovery researcher using AI and molecular simulation. Experienced large-scale researcher. AI-based protein and antibody designer. Top-ranked Kaggle Grandmaster (0.1%) with over 30 AI/ML competition experiences. Certified Machine Learning Expert through Google Developers Experts Program (GDE). Extensive data-driven problem-solving experience.

EXPERIENCE

AI Researcher (AI Drug Discovery Team Leader)

Dec 2021 - Present

Kakao Brain Corp., Seongnam, S. KOREA

- Reproduced and implemented various foundation models for protein.
 - Developed various protein and antibody-specific language models, such as BERT, GPT2, FIM etc.
 - Reproduced various protein-related language models, such as ESM and ProtGPT.
- Conducted cutting-edge research in ML and structural biology to drug discovery and design.
 - Developed flexible protein docking using denoising and equivariant neural networks.
 - Developed reinforcement learning framework for protein sequence design.
 - Developed SOTA inverse-folding model via denoising training.
 - Developed a new type of antibody-antigen complex prediction model.
 - Developed multi-modal foundation model for protein representation learning.
 - Developed a new type of protein representation learning framework focusing on surfaces.
 - Developed a new type of protein generative model focusing on surfaces.
- Managed and conducted antibody design for specific epitopes with experimental scientists.
 - Collaborated with external computational and experimental scientists for de novo antibody design.
 - Designed antibody design project for specific epitopes using various in-house deep learning models (generative models, structure prediction models, affinity prediction models, etc.), and computational tools (Rosetta software).
- Led AI drug discovery team from scratch for two years.
 - Led three doctoral-level researchers.
 - Mentored two master-level intern researchers.

AI Researcher (Large-Scale engineering team)

Oct 2021 – Dec 2021

Kakao Brain Corp., Seongnam, S. KOREA

- Conducted cutting-edge research in multi-modal pre-training.
 - Developed state-of-the-art Multilingual ALIGN models via triple contrastive loss.

AI Research Engineer (Auto-Learn team)

Feb 2021 – Oct 2021

Kakao Brain Corp., Seongnam, S. KOREA

- Reproduced state-of-the-art vision-language large-scale models.
 - Reproduced OpenAI's CLIP and Google's ALIGN.
 - The Reproduced ALIGN is uploaded in HuggingFace Model Card.
- Reproduced various state-of-the-art vision-language models.
 - Reproduced PixelBERT and ViLT for visual question answering research.
- Participated in molecule structure extraction competition using images in documents.

Senior Researcher

Dec 2019 - Jan 2021

Korea Atomic Energy Research Institute, Daejeon, S. KOREA

- Conducted data scientist role.
 - Fulfilled data science duties for the atomic field in a government-funded research institute.
 - Designed and applied deep learning and statistical-based anomaly detection system to multi-variate time series data of nuclear reactor.
 - Developed an object detection system.
- Conducted and lead ML research.
 - Led dataset creation project to generate Korean visual question answering.
 - Developed Korean visual question answering system.

- Conducted research on RNA secondary structure prediction.

EDUCATION

| | |
|--|---------------------|
| Ph.D., Chemical and Biomolecular Engineering, KAIST, S. KOREA | Mar 2016 - Feb 2020 |
| M.S., Chemical and Biomolecular Engineering, KAIST, S. KOREA | Mar 2014 - Feb 2016 |
| B.S., Chemical and Biomolecular Engineering, Pusan National University, S. KOREA | Mar 2008 - Feb 2014 |

PUBLICATIONS (conference)

- [C3] Youhan Lee[†], Hasun Yu[†], Jae-myeong Lee[†], Jaehoon Kim, “Pre-training Sequence, Structure, and Surface Features for Comprehensive Protein Representation Learning”, *NeurIPS 2023 MLSB Workshop* (Machine Learning in Structural Biology), *ICLR 2024 (accepted)*
- [C2] Minjun Kim, SeungWoo Song, Youhan Lee, Haneol Jang, KyungTae Lim, “BOK-VQA: Bilingual Outside Knowledge-based Visual Question Answering via Graph Representation Pretraining”, (accepted) *AAAI 2024* - Thirty-Eighth AAAI Conference on Artificial Intelligence
- [C1] Youhan Lee, KyungTae Lim, Woonhyuk Baek, Byungseok Roh, Saehoon Kim, “Efficient Multilingual Multi-modal Pre-training through Triple Contrastive Loss”, *COLING 2022* - The 29th International Conference on Computational Linguistics

PUBLICATIONS (Journals)

- [J8] Hannah K. Wayment-Steele, Wipapat Kladwang, Andrew M. Watkins, Do Soon Kim, Bojan Tunguz, Walter Reade, Maggie Demkin, Jonathan Romano, Roger Wellington-Oguri, John J. Nicol, Jiayang Gao, Kazuki Onodera, Kazuki Fujikawa, Hanfei Mao, Gilles Vandewiele, Michele Tinti, Bram Steenwinckel, Takuya Ito, Taiga Noumi, Shujun He, Keiichiro Ishi, Youhan Lee, Fatih Öztürk, Anthony Chiu, Emin Öztürk, Karim Amer, Mohamed Fares, Eterna Participants, and Rhiju Das, “Predictive models of RNA degradation through dual crowdsourcing”, *Nature Machine Intelligence* volume 4, pages1174–1184 (2022)
- [J7] Lars A Bratholm, Will Gerrard, Brandon Anderson, Shaojie Bai, Sunghwan Choi, Lam Dang, Pavel Hanchar, Addison Howard, Guillaume Huard, Sanghoon Kim, Zico Kolter, Risi Kondor, Mordechai Kornbluth, Youhan Lee, Youngsoo Lee, Jonathan P Mailoa, Thanh Tu Nguyen, Milos Popovic, Goran Rakocevic, Walter Reade, Wonho Song, Luka Stojanovic, Erik H Thiede, Nebojsa Tijanic, Andres Torrubia, Devin Willmott, Craig P Butts, David R Glowacki, “A Community-powered Search of Machine Learning Strategy Space to Find NMR Property Prediction Models”, *PLOS ONE* 2021, 16(7): e0253612.
- [J6] Byoungil Jeon, Youhan Lee, Myungkook Moon, Jongyul Kim, Gyuseong Cho, “Reconstruction of Compton Edges in Plastic Gamma Spectra Using Deep Autoencoder”, *Sensors* 2020, 20, (10), 2895.
- [J5] Soo-Yeon Cho[†], Youhan Lee[†], Sangwon Lee, Hohyung Kang, Jaehoon Kim, Junghoon Choi, Jin Ryu, Heeun Joo, Hee-Tae Jung, Jihan Kim. (2020). Finding hidden signals in chemical sensors using deep learning”, *Analytical Chemistry*, 92(9), 6529-6537. **Highlighted in** "Neural Network Measures Gas Below a Sensor's Limit", C&EN, May 6th, 2020.
- [J4] Youhan Lee, Roberta Poloni, Jihan Kim, “Probing gas adsorption in MOFs using an efficient ab initio widom insertion Monte Carlo method”, *J. Comput. Chem.*, 2016, 37 (32), 2808-2815.
- [J3] Soo-Yeon Cho, Youhan Lee, Hyeong-Jun Koh, Hyunju Jung, Jong-Seon Kim, Hae-Wook Yoo, Jihan Kim, Hee-Tae Jung, “Superior chemical sensing performance of black phosphorus: Comparison with MoS₂ and graphene”, *Adv. Mater.*, 2016, 28 (32), 7020-7028.
- [J2] Juhuk Choi, Youhan Lee, Jihan Kim, Hyunjoon Lee, “Enhancing stability of octahedral PtNi nanoparticles for oxygen reduction reaction by halide treatment”, *J. Power Sources*, 2016, 307, 883-890.
- [J1] Soo-Yeon Cho, Seon Joon Kim, Youhan Lee, Jong-Seon Kim, Woo-Bin Jung, Hae-Wook Yoo, Jihan Kim, Hee-Tae Jung, “Highly enhanced gas adsorption properties in vertically aligned MoS₂ layers”, *ACS nano*, 2015, 9 (9), 9314-9321.

PUBLICATIONS (Preprint)

- [P4] Youhan Lee, Jaehoon Kim, DNDesign: Enhancing Physical Understanding of Protein Inverse Folding Model via Denoising, 2023, <https://www.biorxiv.org/content/10.1101/2023.12.03.567710v1>
- [P3] Youhan Lee, Jaehoon Kim, ShapeProt: Top-down Protein Design with 3D Protein Shape Generative Model, 2023, <https://www.biorxiv.org/content/10.1101/2023.12.03.567710v1>
- [P2] Jaemyung Lee, Kyeongtak Han, Jaehoon Kim, Hasun Yu, Youhan Lee, Solvent: A Framework for Protein Folding, 2023, <https://arxiv.org/abs/2307.04603>
- [P1] Youhan Lee, Hasun Yu, ProtFIM: Fill-in-Middle Protein Sequence Design via Protein Language Models, 2023, <https://arxiv.org/abs/2303.16452>

GITHUB PROJECTS

- [G2] Jaemyung Lee, Kyeongtak Han, Jaehoon Kim, Hasun Yu, Youhan Lee, Solvent: A Framework for Protein Folding, 2023, <https://github.com/kakaobrain/solvent>

- **[G1]** Boogeon Yoon, **Youhan Lee**, Woonhuk Baek, An implementation of ALIGN on COYO-700M dataset, 2023, <https://github.com/kakaobrain/coyo-align>

HONORS

- Kaggle GrandMaster in Kaggle community (2020)
- Google Developers Experts for Machine Learning (2020)
- 9th gold prize, “Bristol-Myers Squibb – Molecular translation” challenge at Kaggle (2021)
- 3rd prize winner, DramaQA challenge at ECCV 2020 (2020)
- 9th gold prize, “OpenVaccine: COVID-19 mRNA Vaccine Degradation Prediction” challenge at Kaggle (2020)
- 3rd prize winner, “Predicting chemical properties” challenge at Kaggle (2019)
- 3rd prize winner, “APTOS 2019 Blindness Detection” challenge at Kaggle (2019)

SKILLS

- AI-related: TensorFlow, Keras, Pytorch, DGL, Pytorch-geometric, Pytorch-lightning
- Computation-chemistry-related: Rosetta, QuantumEspresso, LAMMPS, COTA
- Programming-related: Git, GitLab, Bash/Shell, Python, C
- Cloud-related: GCP, Azure, AWS

PROFESSIONAL SERVICES

- Reviewer: ACL 2024, GEMBio workshop at ICLR 2024