

# JINHO CHOI

Munich, Germany | jjho.choi@gmail.com | sites.google.com/view/jjhochoi | github.com/jjho-choi | linkedin.com/in/jinhochoi94

## SUMMARY

Ph.D. candidate at KAIST and visiting researcher at Helmholtz Munich, working toward **safe and trustworthy AI** through **mechanistic interpretability** (making the concepts inside models explicit and reproducible) and **data-centric AI** (tracing model behavior back to its training data). Before the Ph.D., five years as an AI scientist at two startups, taking models from research prototype to deployed product: fairness in hiring models (150+ companies), conversational agents (500K+ users), and biomedical imaging.

## EDUCATION

- **Ph.D. in Artificial Intelligence**, KAIST (Advisor: Jaegul Choo), Aug. 2022 – Feb. 2028 expected, part-time 2022–2024
- **M.S. in Computer Science and Engineering**, Korea University — Advisor: Jaegul Choo, Mar. 2017 – Aug. 2019
- **B.S. in Computer Science and Engineering**, Korea University, Mar. 2013 – Feb. 2017

## EXPERIENCE

### Helmholtz Munich

Feb. 2026 - Present

Visiting Researcher, Dynamical Inference Lab (PI: Steffen Schneider), joint project with KAIST DAVIAN Lab Munich, Germany

- Developed **ConSDL**, which extracts the concepts a model has learned reproducibly. With sparse autoencoders, the standard tool, two training runs can disagree on what the concepts are, undermining any analysis built on them. ConSDL recovers the same concepts run after run and is the most consistent method on vision and language models [11, ICLR 2027 under review].

### KAIST, DAVIAN Lab

Aug. 2022 - Present

Ph.D. Researcher

Daejeon, Korea

- Developed **ConceptScope**, which breaks a dataset into human-interpretable visual concepts and measures how each concept distributes across classes, uncovering many previously unreported biases in real-world image datasets [9, NeurIPS 2025].
- Trained sparse autoencoders on the CLIP vision transformer (**PatchSAE**), discovering 49K localized visual concepts that explain how the model reaches its predictions and how that changes under adaptation to new tasks [8, ICLR 2025]. Extended the analysis to large vision–language models, grounding visual concepts at inference time (**VisualScratchpad**) [10, ICLR 2026 Workshop].

### Genesislabs

Sep. 2022 - Nov. 2024

AI Scientist

Seoul, Korea

- Built the **fairness pipeline** for **viewinterHR**, a video interview assessment product used by 150+ companies: a diagnosis stage that quantifies how sensitive and nuisance attributes affect outcomes, and a mitigation algorithm that reduces group disparity in features and outputs, cutting gender bias by 35% with no loss in accuracy [6, IEEE Access 2023].
- Developed a **talking-face generation** method with accurate lip-sync and high video clarity, deployed as the AI interviewer avatar in **viewinterHR**. Worked with HR experts to ensure the generated interviewer reflected real interview dynamics.
- Built the **conversational agents** behind **Zuicy**, LLM personas of YouTube creators (500K+ downloads), using prompt chaining, retrieval-augmented generation, and memory recall, plus an automated pipeline for building each creator persona.

### Tomocube

Feb. 2019 - Sep. 2022

AI Researcher

Seoul, Korea

- Developed **3D cell segmentation** models for four organelle types and cell instances from label-free refractive index images, achieving state-of-the-art accuracy, and integrated them into commercial cell analysis software [4, SPIE 2021] [5, bioRxiv 2021].
- Built a **human-in-the-loop segmentation** model that lifts labeled 2D slices to 3D masks and actively requests correction on uncertain slices, improving annotation speed and accuracy by 9.5% and serving as the in-house annotation system [7, WACV 2024].

## PUBLICATIONS

- [11] **Jinho Choi**, Hyesu Lim, Jaegul Choo, and Steffen Schneider. Beyond SAEs: Consistent Concept Discovery with Sparse Coding. *Under review at the International Conference on Learning Representations (ICLR)*, 2027.
- [10] Hyesu Lim, **Jinho Choi**, Taekyung Kim, Byeongho Heo, Jaegul Choo, and Dongyoon Han. VisualScratchpad: Grounding Visual Concepts in Large Vision Language Models. *ICLR 2026 Workshop on Trustworthy AI*, 2026. [arXiv]
- [9] **Jinho Choi**, Hyesu Lim, Steffen Schneider, and Jaegul Choo. ConceptScope: Characterizing Dataset Bias via Disentangled Visual Concepts. *Conference on Neural Information Processing Systems (NeurIPS)*, 2025. [arXiv] [code] [project page]
- [8] Hyesu Lim, **Jinho Choi**, Jaegul Choo, and Steffen Schneider. Sparse Autoencoders Reveal Selective Remapping of Visual Concepts During Adaptation. *International Conference on Learning Representations (ICLR)*, 2025. [arXiv] [code] [project page]
- [7] Wonwoo Cho, Dongmin Choi, Hyesu Lim, **Jinho Choi**, Saemee Choi, Hyun-seok Min, Sungbin Lim, and Jaegul Choo. Slice and Conquer: A Planar-to-3D Framework for Efficient Interactive Segmentation of Volumetric Images. *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024.
- [6] Changwoo Kim, **Jinho Choi**, Jongyeon Yoon, Daehun Yoo, and Woojin Lee. Fairness-aware Multimodal Learning in Automatic Video Interview Assessment. *IEEE Access*, 11:122677–122693, 2023.
- [5] **Jinho Choi**, Hye-Jin Kim, et al. Label-free Three-dimensional Analyses of Live Cells with Deep-learning-based Segmentation Exploiting Refractive Index Distributions. *bioRxiv*, 2021.

- [4] **Jinho Choi**, Junwoo Park, Hyun-seok Min, Hyungjoo Cho, Sungbin Lim, and Jaegul Choo. 3D Cell Instance Segmentation via Point Proposals using Cellular Components. *SPIE Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XIX*, 2021.
- [3] **Jinho Choi**, Sanghun Jung, Deokgun Park, Jaegul Choo, and Niklas Elmqvist. Visualizing for the Non-Visual: Enabling the Visually Impaired to Use Visualization. *Computer Graphics Forum (EuroVis)*, 38(3):249–260, 2019.
- [2] Minsuk Choi, Sungbok Shin, **Jinho Choi**, et al. TopicOnTiles: Tile-based Spatio-Temporal Event Analytics via Exclusive Topic Modeling on Social Media. *ACM CHI Conference on Human Factors in Computing Systems (CHI)*, 2018.
- [1] Sungbok Shin, Minsuk Choi, **Jinho Choi**, et al. STExNMF: Spatio-Temporally Exclusive Topic Discovery for Anomalous Event Detection. *IEEE International Conference on Data Mining (ICDM)*, 2017.

## SERVICE

---

- **Reviewer:** NeurIPS, ICLR

## SKILLS

---

- **Technical:** Python | PyTorch | Lightning | Hugging Face | Hydra | SLURM / HPC | CUDA | Libtorch | Git | Linux
- **Languages:** Korean (native) | English (professional) | German (elementary, learning)