Joonghyuk Hahn

Email: greghahn@yonsei.ac.kr,
jhhahn0@gmail.com
Homepage: peer0.github.io
GitHub: github.com/peer0
Linkedin: joonghyuk-hahn

Research Vision

My research vision is to build **provably safe, verifiable, and trustworthy AI systems**. I bridge formal language theory and modern AI/NLP, applying theoretical tools—formal grammars and automata—to create AI systems with **mathematical guarantees** on safety and correctness.

- **Core Approach: Bridging Formal Language Theory and Modern AI**
  My Ph.D. research bridges formal language theory and AI/NLP to address this challenge. I advance foundational questions in computational theory—such as decidability of string properties (Parikh matrices), language hierarchies, and formal characterization of structural constraints—and directly apply these theoretical tools to build **verifiable AI systems**. By integrating symbolic, grammar-based methods with neural models, I create frameworks where AI behavior can be formally specified and mathematically verified.

- **Future Focus: Verifiable LLM Safety**
  I aim to apply these methodologies to the most pressing AI challenge: **provably safe LLMs**. By formalizing safety constraints as decidable properties in formal language theory, I will develop systems with **mathematical verification** of outputs, guaranteed abstention mechanisms, and transparent specifications that enable external auditing. My unique expertise in both formal methods and modern NLP positions me to pioneer **verifiable AI safety**—moving beyond heuristics to mathematical guarantees.

Education

Yonsei University — Seoul, KR
Integrated Ph.D. course in Computer Science — 2019–Current
Supervisor: Yo-Sub Han

Thesis: *Symbolic Techniques for Deep Learning in Natural Language Processing and Program Analysis*

Yonsei University — Seoul, KR
Bachelor's degree in Computer Science — 2015–2019

Publications

Full list: [Google Scholar] [DBLP] — † indicates equal contribution

Peer-Reviewed Conference/Journal Papers

[C1] **Query4Regex: Verifiable Regex Transformation through Formal Operations from NL and DSL Queries**
Joonghyuk Hahn, Yo-Sub Han
To appear in Findings of EACL 2026

[C2] **Repairing Regex Vulnerabilities via Localization-Guided Instructions**
Sicheol Sung†, **Joonghyuk Hahn**†, Yo-Sub Han
To appear in Proceedings of EACL 2026

[C3] **A Regex Minimization Benchmark: A PSPACE-Complete Challenge for Language Models**
Hyundong Jin, **Joonghyuk Hahn**, Yo-Sub Han
To appear in Proceedings of EACL 2026

[C4] **EnCur: Curriculum-Based In-Context Learning with Structural Encoding for Code Time Complexity Prediction**
Joonghyuk Hahn, Aditi, Seung-Yeop Baik, Shinwoo Park, Sang-Ki Ko, Yo-Sub Han
To appear in ESWA 2026 [PDF]

[C5] **AmpleHate: Amplifying the Attention for Versatile Implicit Hate Detection**
Yejin Lee, **Joonghyuk Hahn**, Hyeseon Ahn, Yo-Sub Han
In Proceedings of EMNLP 2025 [PDF]

[C6] **CodeComplex: Dataset for Worst-Case Time Complexity Prediction**
Seung-Yeop Baik, **Joonghyuk Hahn**, Jungin Kim, Mingi Jeon, Aditi, Yo-Sub Han, Sang-Ki Ko
In Findings of EMNLP 2025 [PDF]

[C7] **TCProF: Time-Complexity Prediction SSL Framework**
Joonghyuk Hahn, Hyeseon Ahn, Jungin Kim, Soohan Lim, Yo-Sub Han
In Proceedings of NAACL 2025 [PDF]

[C8] **Advanced Code Time Complexity Prediction Approach Using Contrastive Learning**
Shinwoo Park, **Joonghyuk Hahn**, Elizabeth Orwig, Sang-Ki Ko, Yo-Sub Han
In EAAI 2025, Vol. 151 [PDF]

[C9] **Characterizations of M-equivalence and weak M-relation**
Joonghyuk Hahn, Hyunjoon Cheon, Yo-Sub Han
In IJFCS 2025 [PDF]

[C10] **On the Decidability of Infix Inclusion Problem**
Hyunjoon Cheon†, **Joonghyuk Hahn**†, Yo-Sub Han
In Theory of Computing Systems 2024, Vol. 68, 301–321 [PDF]

[C11] **Universal Rewriting Rules for the Parikh Matrix Injectivity Problem**
Ingyu Baek†, **Joonghyuk Hahn**†, Yo-Sub Han, Kai Salomaa
In Proceedings of DLT 2024, LNCS 14791, 68–81 [PDF]

[C12] **SuperST: Superficial Self-Training for Few-Shot Text Classification**
Ju-Hyoung Lee†, **Joonghyuk Hahn**†, Hyeon-Tae Seo, Jiho Park, Yo-Sub Han
In Proceedings of LREC-COLING 2024 [PDF]

[C13] **ATHENA: Mathematical Reasoning with Thought Expansion**
J.B. Kim, Hazel Kim, **Joonghyuk Hahn**, Yo-Sub Han
In Proceedings of EMNLP 2023, 16315–16327 [PDF]

[C14] **GDA: Grammar-based Data Augmentation for Text Classification using Slot Information**
Joonghyuk Hahn, Hyunjoon Cheon, Elizabeth G. Orwig, Su-Hyeon Kim, Sang-Ki Ko, Yo-Sub Han
In Findings of EMNLP 2023 [PDF]

[C15] **M-equivalence of Parikh Matrix over a Ternary Alphabet**
Joonghyuk Hahn, Hyunjoon Cheon, Yo-Sub Han
In Proceedings of CIAA 2023 [PDF]

[C16] **Boosting Code Summarization by Embedding Code Structures**
Jikyeong Son†, **Joonghyuk Hahn**†, Hyeon-Tae Seo, Yo-Sub Han
In Proceedings of COLING 2022, 5966–5977 [PDF]

[C17] **On the Decidability of Infix Inclusion Problem**
Hyunjoon Cheon†, **Joonghyuk Hahn**†, Yo-Sub Han
In Proceedings of DLT 2022, LNCS 13257, 115-126 [PDF]

[C18] **Self-Training using Rules of Grammar for Few-Shot NLU**
Joonghyuk Hahn, Hyunjoon Cheon, Kyuyeol Han, Cheongjae Lee, Junseok Kim, Yo-Sub Han
In Findings of EMNLP 2021 [PDF]

[C19] **Most Pseudo-copy Languages Are Not Context-Free**
Hyunjoon Cheon†, **Joonghyuk Hahn**†, Yo-Sub Han, Sang-Ki Ko
In Proceedings of COCOON 2021, 189–200 [PDF]

Under Review

- **MEC^{3}O: Multi-Expert Consensus for Code Time Complexity Prediction**
  Joonghyuk Hahn†, Soohan Lim†, Yo-Sub Han
  Under review

- **ECO: Enhanced Code Optimization via Performance-Aware Prompting for Code-LLMs**
  Su-Hyeon Kim, Joonghyuk Hahn, Sooyung Cha, Yo-Sub Han
  Under review
- **A Voronoi-Embedding Framework for Semi-Supervised Time-Complexity Prediction**
  Hyeseon Ahn, Joonghyuk Hahn, Yo-Sub Han
  Under review

Projects

2023–2025: Human-AI Programming Platform Research (follow-up)

Ministry of Science and ICT

- Efficient Frameworks for programming platform with human and AI collaboration.

2022–2022: 5th AI Grand Challenge

Ministry of Science and ICT

- Developed a deep learning model for solving mathematical equation problems in a three-stage national-level competition.
- Passed the first stage and wrote the proposal for the competition projects. Then, participated in the second-stage competition.

2019–2022: Malware Pattern Extraction & Human-AI Programming Platform Research

Ministry of Science and ICT

- Researched synergies between NLP and source code, such as code summarization and complexity analysis.

Research Interests for Future Works

- **Mechanistic Interpretability of Generative Models**: Applying formal language theory and automata to understand internal mechanisms of LLMs—analyzing learned circuits, feature representations, and reasoning processes to make models more transparent and interpretable.
- **Formal Reasoning & Symbolic Integration**: Bridging symbolic methods (formal grammars, logical constraints, decidability analysis) with neural models to enable verifiable reasoning, compositional generalization, and systematic problem-solving in language models.
- **Constrained & Controllable Generation**: Developing grammar-based frameworks for structured, safe generation with guarantees on output properties—applicable to code synthesis, mathematical reasoning, and domain-specific language generation.
- **Efficient & Robust Language Models**: Investigating byte-level models, robustness to perturbations, and efficient architectures informed by formal language hierarchies—advancing foundations for more resilient and efficient generative systems.
- **Neuro-Symbolic AI Safety**: Bridging symbolic reasoning (formal methods, logical constraints) with neural models to create interpretable, verifiable AI systems with transparent failure modes.

Technical Skills

- **Programming Languages**: Python, C/C++, Java
- **ML/NLP Frameworks**: PyTorch, TensorFlow, Hugging Face Transformers
- **Formal Methods**: Automata theory, formal grammars (CFG, PDA), decidability analysis, formal language hierarchies
- **Research Areas**: Mechanistic interpretability, constrained generation, neuro-symbolic AI, formal verification for neural networks
- **Tools**: LaTeX, Git, Linux/Unix

References

Prof. Yo-Sub Han Yonsei University
Email: emmous@yonsei.ac.kr

- Supervisor, Department of Computer Science

Prof. Sang-Ki Ko University of Seoul
Email: sangkiko@uos.ac.kr

- Department of Artificial Intelligence

Prof. Sang-Min Choi Gyeongsang National University
Email: jerassi@gnu.ac.kr

- Department of Computer Science