

VÉSTEINN SNÆBJARNARSON

+354-66 33 727 | vest.snae@gmail.com | vesteinn.is | github.com/vesteinn

EDUCATION

University of Copenhagen, Pioneer Centre for Artificial Intelligence <i>PhD Fellow in Computer Science (ELLIS)</i> <ul style="list-style-type: none">Advised by Serge Belongie and Ryan Cotterell at ETH Zürich.Thesis on algorithmic transformations of language models and robust tokenization.	Copenhagen, Denmark <i>Autumn 2022 -</i>
University of Iceland <i>MS in Computer Science - First Class with Distinction</i>	Reykjavík, Iceland <i>October 2021</i>
University of Iceland <i>BA in Philosophy - First Class</i>	Reykjavík, Iceland <i>June 2012</i>
University of Iceland <i>BS in Mathematics - First Class</i>	Reykjavík, Iceland <i>October 2011</i>

EXPERIENCE

Visiting Researcher <i>University of Iceland</i> <ul style="list-style-type: none">Visiting the Computer Science department.	September 2025 – August 2026 <i>Reykjavík, Iceland</i>
Visiting Researcher <i>ETH Zürich</i> <ul style="list-style-type: none">One year stay at Rycolab.Worked on transformation of language models, causal evaluation of language models, among other things.	September 2024 – August 2025 <i>Zürich, Switzerland</i>
NLP Researcher / CTO <i>Miðeind ehf - Language technology</i> <ul style="list-style-type: none">Currently part-time consulting on TrustLLM Horizon Europe project.Promoted to CTO in autumn 2021.General purpose and customized machine translation systems.Training and analysis of language models, monolingual and cross-lingual.Development of various core NLP tools for Icelandic.	November 2019 – February 2023 <i>Reykjavík, Iceland</i>
Senior Developer <i>Trackwell ehf</i> <ul style="list-style-type: none">Full-stack web development for fleet- and human-resource management solutions.Modernised systems. Front-ends written in react and containerization of back-end systems.Designed multi-tenancy solution for separated clients that had previously been running on multiple servers.	January 2018 – October 2019 <i>Reykjavík, Iceland</i>
Developer / PM / CTO <i>Cooori ehf - Icelandic / Japanese ed-tech company</i> <ul style="list-style-type: none">CTO from 2015.Ed-tech company. Pivoted from teaching Japanese to a global audience to teaching English in Japan.	August 2012 – March 2017 <i>Tokyo, Japan</i>
Research Assistant <i>University of Iceland, Faculty of Life and Environmental Sciences</i> <ul style="list-style-type: none">Assembled and analyzed genomic, transcriptomic, and epigenetic data, mainly in lichens.Developed methods for handling assembly of metagenomic samples using multiple sequencing types.	February 2010 – August 2012 <i>Reykjavík, Iceland</i>

TEACHING

Philosophy of Computation <i>Head TA, ETH Zürich</i>	Spring 2025
Large Language Models <i>TA, ETH Zürich</i>	Spring 2025
Advanced Topics in Machine Learning <i>TA, ETH Zürich</i>	Autumn 2024
Vision and Image Processing <i>TA, University of Copenhagen</i>	Autumn 2023
Mathematical Analysis and Probability Theory <i>TA, University of Copenhagen</i>	Autumn 2023
Introduction to Language Technology <i>Head Teacher, University of Iceland</i>	Autumn 2021
<ul style="list-style-type: none">• Masters-level course in the Computer Science department.• Focus on NLP using deep learning and classical methods.• Designed the course, heavily inspired by CS224n at Stanford.	
AI lifecycle <i>Guest lecturer, University of Iceland</i>	Spring 2021
<ul style="list-style-type: none">• Lecture on neural machine translation with a case study.	
Introduction to Logic <i>TA, University of Iceland</i>	Spring 2012
Mathematical Analysis <i>TA, University of Iceland</i>	Autumn 2009

TALKS

Looking Back on The Icelandic Language Technology Program <i>ASTIN '24</i>	November 2024
<ul style="list-style-type: none">• Talk about the government-funded program and its wide reaching implications.	
How correct is your VLM? <i>Visipedia Workshop '24</i>	April 2024
<ul style="list-style-type: none">• Evaluation of vision-and-language models in the context of the iNaturalist dataset.	
Language Models and Innovation <i>University of Iceland Innovation Award Ceremony</i>	May 2023
<ul style="list-style-type: none">• Impact of artificial intelligence on innovation.	
Representation Learning and Neural Network <i>Arion Bank</i>	March 2022
<ul style="list-style-type: none">• Representation learning and AI, presented for one of the largest banks in Iceland.	
IceBERT – language models for Icelandic <i>Icelandic Linguistics Society</i>	November 2021
<ul style="list-style-type: none">• Language models and NLP for Icelandic.	
Artificial Intelligence and Question Answering <i>Icelandic Computer Science Society</i>	November 2021
<ul style="list-style-type: none">• Lecture on AI and QA based on MSc thesis.	
Language and Artificial Intelligence <i>Reykjavík City library</i>	September 2021
<ul style="list-style-type: none">• An informal conversation for the public with a journalist interviewing. General discussion about AI and language.	
Machine Translation, an Important Option <i>Reykjavík University</i>	May 2021
<ul style="list-style-type: none">• Machine translation for Icelandic and its applications. Part of a conference about language technology, Máltækniþytingin (e. <i>The Language Technology Revolution</i>)	

PUBLICATIONS

- Vésteinn Snæbjarnarson**, Samuel Kiegele, Tianyu Liu, Reda Boumasmoud, Tim Viera, Ryan Cotterell. *Transducing Language Models*, Salon des Refusés at EurIPS (non-archival), 2025.
- Jannek Ulm, Kevin Du, **Vésteinn Snæbjarnarson**. *Contrastive Decoding for Synthetic Data Generation in Low-Resource Language Modeling*. BabyLM workshop, EMNLP 2025.
- Vésteinn Snæbjarnarson**, Kevin Du, Niklas Stoehr, Ryan Cotterell, Serge Belongie, Nico Lang, Stella Frank. *Taxonomy-Aware Evaluation of Vision-Language Models*. CVPR 2025.
- Shauli Ravfogel, Anej Svete, **Vésteinn Snæbjarnarson** and Ryan Cotterell. *Gumbel Counterfactual Generation From Language Models*. ICLR 2025.
- Sebastian Bugge Loeschke, Mads Toftrup, Michael Kastoryano, Serge Belongie and **Vésteinn Snæbjarnarson**. *LoQT: Low Rank Adapters for Quantized Pre-training*. NeurIPS 2024 & WANT@ICML 2024 Oral. (Best paper award.)

Niklas Stoehr, Kevin Du, **Vésteinn Snæbjarnarson**, Robert West, Ryan Cotterell and Aaron Schein. *Activation Scaling for Attribution and Intervention in Language Models*. EMNLP 2024.

Evgenia Ilyinskaya, **Vésteinn Snæbjarnarson**, Hanne Krage Carlsen, and Björn Oddsson. *Small-scale geohazards cause significant and highly variable impacts on emotions*. 2024. Natural Hazards and Earth System Sciences (NHES).

Kevin Du, **Vésteinn Snæbjarnarson**, Niklas Stoehr, Jennifer C. White, Aaron Schein, Ryan Cotterell. *Persuasiveness and Susceptibility in In-Context Learning*. Annual Meeting of the Association for Computational Linguistics (ACL) 2024.

Steinunn Rut Friðriksdóttir, Annika Simonsen, Atli Snær Ásmundsson, Guðrún Lilja Friðjónsdóttir, Anton Karl Ingason, **Vésteinn Snæbjarnarson** and Hafsteinn Einarsson. *Ice and Fire: Dataset on Sentiment, Emotions, Toxicity, Sarcasm, Hate speech, Sympathy and More in Icelandic Blog Comments*. TRAC 2024; LREC-COLING. Torino, Italy.

Nico Lang, **Vésteinn Snæbjarnarson**, Elijah Cole, Oisín Mac Aodha, Christian Igel, Serge Belongie. *From Coarse to Fine-Grained Open-Set Recognition*. Conference on Computer Vision and Pattern Recognition (CVPR) 2024. Seattle, USA.

Peter Ebert Christensen, **Vésteinn Snæbjarnarson**, Andrea Dittadi, Serge Belongie and Sagie Benaim. *Assessing Neural Network Robustness via Adversarial Pivotal Tuning*. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV, oral) 2024. Hawaii, USA.

Idan Schwartz*, **Vésteinn Snæbjarnarson***, Sagie Benaim, Hila Chefer, Lior Wolf, Serge Belongie. 2023. *Discriminative Class Tokens for Text-to-Image Diffusion Models*. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) 2023. Paris, France.

Svanhvít Lilja Ingólfssdóttir, Petur Orri Ragnarsson, Haukur Páll Jónsson, Haukur Barri Simonarson, Vilhjalmur Þorsteinsson and **Vésteinn Snæbjarnarson**. *Byte-Level Grammatical Error Correction Using Synthetic and Curated Corpora*. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023, oral). Toronto, Canada.

Vésteinn Snæbjarnarson, Annika Simonsen, Goran Glavaš, Ivan Vulić. *Transfer to a Low-Resource Language via Close Relatives: The Case Study on Faroese*. The 24th Nordic Conference on Computational Linguistics, 2023. Tórshavn, Faroe Islands.

Vésteinn Snæbjarnarson and Hafsteinn Einarsson. 2022. *Cross-Lingual QA as a Stepping Stone for Monolingual Open QA in Icelandic*. In Proceedings of the Workshop on Multilingual Information Access (MIA) at NAACL 2022. Seattle, WA, USA.

Vésteinn Snæbjarnarson and Hafsteinn Einarsson. 2022. *Natural Questions in Icelandic*. In Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC 2022). Marseille, France.

Vésteinn Snæbjarnarson*, Haukur Barri Símonarson*, Pétur Orri Ragnarsson, Svanhvít Lilja Ingólfssdóttir, Haukur Páll Jónsson, Vilhjalmur Þorsteinsson and Hafsteinn Einarsson. 2022. *A Warm Start and a Clean Crawled Corpus – A Recipe for Good Language Models*. In Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC 2022). Marseille, France.

Anna Björk Nikulásdóttir, Þórunn Arnardóttir, Jón Guðnason, Þorsteinn Daði Gunnarsson, Anton Karl Ingason, Haukur Páll Jónsson, Hrafn Loftsson, Hulda Óladóttir, Einar Freyr Sigurðsson, Atli Þór Sigurgeirsson, **Vésteinn Snæbjarnarson** and Steinþór Steingrímsson. 2021. *Help Yourself from the Buffet: National Language Technology Infrastructure Initiative on CLARIN-IS*. In Proceedings of the CLARIN Annual Conference.

Haukur Barri Símonarson, **Vésteinn Snæbjarnarson**, Pétur Orri Ragnarsson, Haukur Páll Jónsson and Vilhjalmur Þorsteinsson. 2021. *Miðeind's WMT 2021 Submission*. In Proceedings of the Sixth Conference on Machine Translation (WMT21, EMNLP), Dominican Republic.

Haukur Páll Jónsson*, Haukur Barri Símonarson*, **Vésteinn Snæbjarnason***, Steinþór Steingrímsson* and Hrafn Loftsson. 2020. *Experimenting with Different Machine Translation Models in Medium-Resource Settings*. In Proceedings of the 23rd International Conference on Text, Speech and Dialogue (TSD 2020). Brno, Czech Republic (Online). Lecture Notes in Computer Science, vol 12284. Springer, Cham.

Vivian P. W. Miao, Sheeba S Manoharan-Basil, **Vésteinn Snæbjarnarson** and Ólafur Andr sson. 2012. *Expression of lec-1, a mycobiont gene encoding a galectin-like protein in the lichen Peltigera membranacea*. Symbiosis 57, 23-31.

OTHER

- GenLM** 2025 -
- Active within the GenLM initiative on constrained decoding.
- SwissAI Multimodal Project** 2024
- Involved in the multimodal horizontal within the SwissAI project.
- TrustLLM - Horizon Europe** 2023-2024
- Joint project with multiple European institutions to develop an open, trustworthy, Germanic LLM.
 - Co-lead in work package on alignment, involved in other packages.
- Icelandic Language Technology Transfer Fund (Mark a tlun   tungu og t ekni)** 2022-2023
- Grant for developing a system capable of automated captioning of English videos in Icelandic. Collaboration with the three largest local distributors of TV material, including the national provider.
 - Another grant for developing an Icelandic-Polish translation system for the City of Reykjav k.
- The University of Iceland Science and Innovation Motivational Award** Summer 2021
- For MSc thesis. Awarded annually to a project affiliated with the University in some way.
- Icelandic Student Innovation fund** Spring 2020
- Grant for creating an Icelandic QA corpus - Natural Questions in Icelandic.
 - Employed five students in the summer of 2020 to create the corpus.
- Icelandic Student Innovation fund** Summer 2010
- Developed methods for posthoc assembly of genomic data from different sequencing methods.

ACADEMIC SERVICE

Reviewing: CVPR (2026), NeurIPS (2025, 2024), ICML (2025), ICLR (2024, 2025), ACL (2025,2024,2023), COLING (2022), EMNLP (2024,2025), Bioinformatics (2012).

Chairing: Co-chair at *Machine Learning and Statistical Methods* session at LREC 2022.

Supervision: Christoffer Ringgaard (BSc), 2023. Jannek Ulm, ETH Z rich, 2025 (MSc).

Thesis examiner: Automatic Abstractive Summarization for Icelandic Text, 60 ECTS, P r Sverisson, University of Iceland, 2023.