NORDIC MACHINE INTELLIGENCE

Nordic Machine Intelligence, vol. 5, pp. 1–14, 2025
Received 01 Aug 2025 / published 05 Aug 2025
https://doi.org/10.5617/nmi.12551

Proceedings of NORA's annual conference 2025

Geir Halnes[1], Anam Javaid[1], Michael Solvang[1], Stefano Nichele[2,3], Michael A. Riegler[4,5], Pankaj Pandey[6], Jacob Hay[7], Hamze Issa[7], Daniele Fantin[7], David Parkes[8], Amber Leeson[8], Malcom McMillan[8], Kate Briggs[8], Jan Wuite[9], Thomas Nagler[9], Kushtrim Visoka[1], Andrea Alessandro Gasparini[1,10], Lina Plataniti[11], Synnøve Rubach[12], Kari-Anne Lyng[11], Preben Castberg[10], Klaus Johannsen[13], Xue-Cheng Tai[13], Gro Fonnes[13], Junyong You[13], Arangan Subramaniam[14], Changkyu Choi[15], Nils Olav Handegard[16], Robert Jenssen[15], Ali Ramezani-Kebrya[10], Helge Fredriksen[17], Felix S. Reimers[2], Ola Huse Ramstad[3], Axel Sandvig[18], Ioanna Sandvig[18], Christopher Michael Skeide Vibe[2], Mikkel Lepperød[19], Solve Sæbø[20], Hao Liu[21], Raymond H. Chan[22], Lingfeng Li[23], Aaron de Leyos[24], Alexander Johannes Stasik[24,25], Signe Riemer-Sørensen[25], Matteo Iervasi[26], Florenc Demrozi[26], Arezo Shakeri[26], Mina Farmanbar[26], Julia Kropiunig[27], Øystein Sørensen[27], Bjørn-Jostein Singstad[28,29,30], Semra Oztemel Sari[31], Mathis Korseberg Stokke[31,32,33], Arina Surko[2], Hasan Oğul[2], Sinan Uğur Umu[34], Martin T. Horsch[35], Fadi Al Machot[35], Maria Bashir[35], Heinz A. Preisig[36], Shailendra Singh[35], Mehdi HoushmandSarkhoosh[3,37], Cise Midoglu[37], Saeed Shafiee Sabet[37], Tomas Kupka[37], Pål Halvorsen[3,37,5], Kjetil Indrehus[10], Ali Ramezanikebrya[10], Aslak Djupskås[24], Ryan Marinelli[10], Anton Tkachenko[38], Benjamin Adolphi[38], Ibrahim Riza Hallac[2], Abdelaziz Qassi[2], Anja Stein[39], Waldir Leoncio Netto[40], David S. Leslie[39], Shakiba Sadat Mirbagheri[41], Claudio Sartori[41], Mehrzad Abdi Khalife[42], Luis M. Lopez-Ramos[5], Emilio Ruiz-Moreno[43], Baltasar Beferull-Lozano[43]

[1]Norwegian Artificial Intelligence Research Consortium (NORA), Norway, [2]Department of Computer Science and Communication, Østfold University College, Halden, Norway, [3]Department of Computer Science, Oslo Metropolitan University, Oslo, Norway, [4]Department of Cyber Security, Simula Research Laboratory, Oslo, Norway, [5]Department of Holistic Systems, Simula Metropolitan Center for Digital Engineering, Oslo, Norway, [6]Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Gjøvik, Norway, [7]Science and Technology AS, Oslo, Norway, [8]Lancaster Environment Centre, Lancaster University, Lancaster, UK, [9]ENVEO IT GmbH, Innsbruck, Austria, [10]Department of Informatics, University of Oslo, Oslo, Norway, [11]NORSUS AS, Fredrikstad, Norway, [12]Becour AS, Fredrikstad, Norway, [13]NORCE Norwegian Research Centre AS, Bergen, Norway, [14]Department of Physics, University of Oslo, Oslo, Norway, [15]Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway, [16]Norwegian Institute of Marine Research, Bergen, Norway, [17]Department of Computer Science, UiT The Arctic University of Norway, Bodø, Norway, [18]Department of Neuromedicine and Movement Science, NTNU, Trondheim, Norway, [19]Department of Numerical Analysis and Scientific Computing, Simula Research Laboratory, Oslo, Norway, [20]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway, [21]Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong SAR, [22]Lingnan University, Tuen Mun, Hong Kong SAR, [23]Hong Kong Centre for Cerebro-Cardiovascular Health Engineering, Hong Kong SAR, [24]Department of Data Science, NMBU, Ås, Norway, [25]Department of Mathematics and Cybernetics, SINTEF Digital, Oslo, Norway, [26]Department of Electrical Engineering and Computer Science, University of Stavanger, Stavanger, Norway, [27]LCBC, Department of Psychology, University of Oslo, Oslo, Norway, [28]Research and Innovation Department, Vestfold Hospital Trust, Tønsberg, Norway, [29]Akershus University Hospital (Medical Technology and E-Health, Lørenskog, Norway, [30]Institute of Clinical Medicine, University of Oslo, Oslo, Norway, [31]Institute for Experimental Medical Research, Oslo University Hospital and University of Oslo, Oslo, Norway, [32]KG Jebsen Centre for Cardiac Research, University of Oslo, Oslo, Norway, [33]Department of Cardiology, Oslo University Hospital, Rikshospitalet, Oslo, Norway, [34]Department of Pathology, University of Oslo, Oslo, Norway, [35]Material Theory and Informatics Group, Norwegian University of Life Sciences (NMBU), Ås, Norway, [36]Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, [37]Forzasys AS, Oslo, Norway, [38]Security Research Department, Promon, Oslo, Norway, [39]School of Mathematical Sciences, Lancaster University, Lancaster, UK, [40]Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Oslo, Norway, [41]Department of Computer Science and Engineering, University of Bologna, Bologna, Italy, [42]Ticker Bell Kft. Company, Budapest, Hungary, [43]Signal and Information Processing for Intelligent Systems,Simula Metropolitan Center for Digital Engineering, Oslo, Norway

Abstract

NORA - The Norwegian Artificial Intelligence Research Consortium - works to strengthen Norwegian research, education and innovation within the fields of AI, machine learning and robotics. NORA is a collaboration between 8 universities, 5 university colleges and 4 research institutes. The current arti-

all three models.

We discovered that the performance of Monte Carlo Dropout is highly dependent on the architectural choices and training configurations. Lastly, MCD results in unjustified high confidence in both interpolation and extrapolation, unlike the GP and BNN.

23. SAIL-K Framework For Secure AI Applications.

Ryan Marinelli[1].
[1]Department of Informatics, University of Oslo, Oslo, Norway.

In this work, a new framework, SAIL-K for engaging with AI systems is proposed. The goal of this framework is to augment social blueprints to orient the focus of developers to create more robust and secure AI systems. Key issues are identified per layer with a sketch of solutions to prime intervention.

24. Deconstructing Obfuscation: A Four-Dimensional Framework for LLMs in Assembly Code Deobfuscation.

Anton Tkachenko[1], Dmitrij Suskevic[1], Benjamin Adolphi[1].
[1]Security Research Department, Promon, Oslo, Norway.

This research systematically evaluates eight state-of-the-art commercial Large Language Models (LLMs) on their ability to analyze and deobfuscate assembly code from obfuscated binaries. Using a known C program obfuscated with Obfuscator-LLVM, we tested GPT-3o Mini High, GPT-4o, GPT-4.5, O1 Pro Mode, DeepSeekR1, Grok3, Grok2, and Claude 3.7 Sonnet against four obfuscation techniques: bogus control flow, instruction substitution, control flow flattening, and combined techniques. We quantified human intervention required for successful deobfuscation on a six-level scale, from fully autonomous (Level 0) to beyond expert correction (Level 5).

Our findings reveal significant performance variations across techniques and models. While several LLMs successfully deobfuscated individual techniques with minimal guidance, all models universally failed against combined obfuscation. We introduce a novel four-dimensional theoretical framework—Reasoning Depth, Pattern Recognition, Noise Filtering, and Context Integration—that systematically explains these variations and pinpoints specific capability limitations in current AI systems.

We identified and classified five recurring error patterns: predicate misinterpretation, structural mapping errors, control flow misinterpretation, arithmetic transformation errors, and constant propagation errors. These patterns reveal fundamental limitations in how LLMs process obfuscated code, particularly when mathematical reasoning and complex pattern recognition are required simultaneously.

This research has dual implications for cybersecurity: it highlights significant vulnerabilities in legitimate software protection mechanisms while informing potential defensive applications in malware analysis. Our empirically-derived three-tier resistance model provides actionable insights for developing both more resilient obfuscation techniques and improved automated analysis tools in an increasingly AI-augmented cybersecurity landscape.

25. Detecting AI Influence in Student Writing: Toward Reliable and Interpretable Classifiers.

Ibrahim Riza Hallac[1], Abdelaziz Qassi[1], Hasan Ogul[1].
[1]Department of Computer Science and Communication (Østfold University College, Halden 1757, Norway).

Large language models are now widely used in student writing, raising new challenges for academic integrity, authorship, and fairness. This work develops reliable classifiers to detect AI influence in English-language student essays, distinguishing between texts written fully by students, edited with AI, or generated mostly by AI. We augment publicly available student writing datasets with no known AI involvement by generating AI-edited and AI-generated versions using controlled prompting with large language models. This process yields an augmented dataset for detection studies.

We train transformer-based language models for three-way classification and compare their accuracy and inference-time efficiency with open-source baselines and commercial detectors such as GPTZero. Teachers review a small set of challenging cases to enhance model interpretability and guide alignment. Active learning is used to prioritize uncertain predictions for human feedback, and preference-based tuning helps adjust model outputs to reflect teacher judgments. This work contributes an augmented dataset for detection studies and a comparative evaluation of detection strategies, offering insights into detection effectiveness, model generalization, and the trade-offs involved in educational deployment.

26. Sequential Preference Learning with the Bayesian Mallows Model.

Øystein Sørensen[1], Anja Stein[2], Waldir Leoncio Netto[3], David S. Leslie[2].
[1]Department of Psychology, University of Oslo, Oslo, Norway.
[2]School of Mathematical Sciences, Lancaster University, Lancaster, United Kingdom.
[3]Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Oslo, Norway.

Data in the form of rankings or preferences, such as click data, consumer preferences, social hierarchies, and voting records, are common across various applications, and are challenging to analyze due to their discrete nature.

The Bayesian Mallows model is a flexible tool for analyzing data in the form of complete or partial rankings,