School of Computing, Engineering, and the Built Environment Edinburgh Napier University

PHD STUDENT PROJECT

Application instructions:
Detailed instructions are available at :
https://www.napier.ac.uk/research-and-innovation/doctoral-college/how-to-apply

Prospective candidates are encouraged to contact the Director of Studies (see details below) to discuss the project and their suitability for it.

Project details

Supervisory Team:

- Director of Study: Dr Peter Barclay (Email: p.barclay@napier.ac.uk)
- 2ND Supervisor: tbc

Subject Group: Computer Science

Research Areas: Computer Science - Machine Learning

Project Title: Identifying artificially generated creative writing

Project description:

Large Language Models (LLM) have found many uses, such as chatbots for customer support or helping to debug code. However, these models have also been misused, for example by generating fake news stories to spread misinformation. The ability to detect machine-generated content would help address the harm caused by the misuse of LLMs.

Prior research has focused on the identification of deceptive text in a variety of areas, including phishing attempts, fake product review, and academic plagiarism. There is, however, little research on identifying other forms of generated text such as automatic translations, or text that has been reworded by ‘essay assistant’ software.

Notably, the literature shows that creative writing has rarely been considered, although automatic generation of texts such as poems, songs and novels could cause economic harm to creative artists, and may lead to a reduction in quality of the works available. Therefore, this project will focus on characterising the differences between human generated and machine generated text in the domain of creative writing, and construct a classifier to distinguish reliably between human and AI generated text.

References:

Crothers, E. N., Japkowicz, N., & Viktor, H. L. (2023). Machine-Generated Text: A - Comprehensive Survey of Threat Models and Detection Methods. IEEE Access, 11, 70977–71002. https://doi.org/10.1109/ACCESS.2023.3294090

Dhaini, M., Poelman, W., & Erdogan, E. (2023). Detecting ChatGPT: A Survey of the State of Detecting ChatGPT-Generated Text (arXiv:2309.07689). arXiv http://arxiv.org/abs/2309.07689

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text Journal for Educational Integrity, 19(1), Article 1. https://doi.org/10.1007/s40979-023-00140-5

Gehrmann, S., Strobelt, H., & Rush, A. M. (2019). GLTR: Statistical Detection and Goyal, R., Kumar, P., & Singh, V. P. (2023). A Systematic survey on automated text generation tools and techniques - Application, evaluation, and challenges. Multimedia Tools and Applications, 82(28), 43089–43144. https://doi.org/10.1007/s11042-023-15224-0

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work. The New York Times: https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html

Islam, N., Sutradhar, D., Noor, H., Raya, J. T., Maisha, M. T., & Farid, D. M. (2023). Distinguishing Human Generated Text From ChatGPT Generated Text Using Machine Learning (arXiv:2306.01761). arXiv. http://arxiv.org/abs/2306.01761

Jawahar, G., Abdul-Mageed, M., & Lakshmanan, L. V. S. (2020). Automatic Detection of Machine Generated Text: A Critical Survey (arXiv:2011.01314). arXiv. http://arxiv.org/abs/2011.01314

OpenAI (2023) New AI classifier for indicating AI-written text.. Retrieved 25 October 2023, from https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text OpenAI Platform. (2024).

Puttarattanamanee, M., Boongasame, L., & Thammarak, K. (2023). A Comparative Study of Sentiment Analysis Methods for Detecting Fake Reviews in E-Commerce.
HighTech and Innovation Journal, 4(2), Article 2.

Robins-Early, N. (2024, April 2). Billie Eilish, Nicki Minaj, Stevie Wonder and more musicians demand protection against AI. The Guardian. https://www.theguardian.com/technology/2024/apr/02/musicians-demand-protection-against-ai

Su, J., Zhuo, T. Y., Wang, D., & Nakov, P. (2023). DetectLLM: Leveraging Log Rank Information for Zero-Shot Detection of Machine-Generated Text (arXiv:2306.05540).
arXiv. http://arxiv.org/abs/2306.05540

Venkatraman, S., Uchendu, A., & Lee, D. (2023). GPT-who: An Information Density-based Machine-Generated Text Detector (arXiv:2310.06202). arXiv. http://arxiv.org/abs/2310.06202

Verma, V., Fleisig, E., Tomlin, N., & Klein, D. (2023). Ghostbuster: Detecting Text Ghostwritten by Large Language Models. arXiv.Org. https://www.proquest.com/docview/2819139768

Walters, W. H. (2023). The Effectiveness of Software Designed to Detect AI-Generated Writing: A Comparison of 16 AI Text Detectors. Open Information Science, 7(1).
https://doi.org/10.1515/opis-2022-0158

Candidate characteristics

Education:

Minimum 2:1 degree - Artificial Intelligence, Computer Science, Statistics

Subject knowledge:

Machine Learning

Essential attributes:

- Self-motivated
- Numerate
- Good command of English
- Programming Experience