Doctoral School of Information and Biomedical Technologies Polish Academy of Sciences (TIB PAN)

SUBJECT:

Methods, tools, and tricks for protecting vulnerable groups against online frauds in the era of GenAI

SUPERVISOR:

- Prof. Adam Wierzbicki (main supervisor);
  Polsko-Japońska Akademia Technik Komputerowych, ul. Koszykowa 86, Warszawa
  adamw@pja.edu.pl

- Radosław Nielek, PhD (auxiliary supervisor).
  Naukowa i Akademicka Sieć Komputerowa – Państwowy Instytut Badawczy (NASK-PIB), ul. Kolska 12, Warszawa
  radoslaw.nielek@nask.pl

Large Language Models (LLMs) and, more generally, generative AI are rapidly changing the landscape of online fraud attempts. Less than 60 seconds of voice samples is enough to train the model to perfectly mimic the tone of voice and style of speaking of any person. Deepfakes are hard to distinguish from real shots even for experts. These technologies, combined with the automatic collection of personal data, cognitive digital twins, and persuasive text generation, create a powerful tool for malicious actors to launch personalized fraud attempts. Decreasing costs will enable the organization of spear attacks at scale for pennies. Elderly people, people with mild cognitive impairment, low digital competencies, and children will be particularly at risk. Researchers, and NASK in particular, as a national-level Computer Emergency Response Team, are obliged to take the lead in preparing an adequate answer.

Recent studies focus on automatic threat detection with the aid of AI models or designing user interfaces that help minimize cognitive errors. Both approaches are promising but are still at an early stage or prone to a cat-and-mouse game. Moreover, new methods are needed to address opportunities and challenges brought by novel intermediaries such as AR//VR, AI assistants, robots, and, in the near future, brain-computer interfaces.

This PhD research focuses on developing and testing novel tools, procedures, and methods for protecting vulnerable individuals against targeted online fraud attempts. The research question includes, among others, automatically identifying and scoring vulnerable individuals, leveraging cognitive mechanisms and heuristics for protecting users, application of state-of-the-art AI solutions for counteracting state-of-the-art GenAI-driven attacks, and developing and implementing a trustworthy-by-design approach.

The outcomes of this research are intended to have not only a theoretical but also a strong practical impact and are meant to be truly interdisciplinary, combining computer science, cognitive science, and psychology. Extensive use of unique data on human behavior and real events, as well as conducting experiments, is planned.

REQUIREMENTS:

- MSc degree in Psychology, Cognitive studies, Computer Science, Artificial Intelligence, or a related field,
- At least basic programming skills (Python preferred),
- Experience with UX/UI studies and qualitative/quantitative research,
- Knowledge of current research on UX/UI, cognitive biases, and cybersecurity,
- Good practical knowledge of current GenAI tools,
- Advanced level of English (spoken and written);

BIBLIOGRAPHY:

- Nielek, Radosław, et al. "Foraging in Multi-List Recommender Interfaces: the Effects of Digital Nudges and Aging." *International Journal of Human-Computer Studies* (2025): 103588.
- Knowles, Bran, et al. "The harm in conflating aging with accessibility." *Communications of the ACM* 64.7 (2021): 66-71.
- Sharevski, Filipo. "Inclusive involvement of at-risk users in cybersecurity research." *IEEE Security & Privacy* (2024).
- Lazarus, Suleman, Peter Tickner, and Michael R. McGuire. "Cybercrime against senior citizens: Exploring ageism, ideal victimhood, and the pivotal role of socioeconomics." *Security Journal* 38.1 (2025): 1-23.
- Nawi, Haslinda Sutan Ahmad, Siti Fatimah Omar, and Suzana Basaruddin. "Securing the Silver Surfers: Cybersecurity Awareness Among Older Adults." *2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM)*. IEEE, 2025.
- Zhai, Yuxiang, et al. "Hear Us, then Protect Us: Navigating Deepfake Scams and Safeguard Interventions with Older Adults through Participatory Design." *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 2025.
- Wierzbicki, Adam, et al. "Computer-implemented visual decision support method and system for comparing products or services." U.S. Patent Application No. 18/506,850.
- Noah, Naheem, et al. "Evaluating privacy & security of online dating applications with a focus on older adults." *2024 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2024.
- Bong, Way Kiat, and Yuan Jing Li. "How Many Times Do I Need to Say,‘It Is Me’? Investigating Two-Factor Authentication with Older Adults in Norway.", https://www.scitepress.org/Papers/2025/132065/132065.pdf
- Herrera, L. D., London Van Sickle, and Ashley Podhradsky. "Bridging the Protection Gap: Innovative Approaches to Shield Older Adults from AI-Enhanced Scams." *2024 Cyber Awareness and Research Symposium (CARS)*. IEEE, 2024.
- Tamut, Hayin, and Indira Kalyan Dutta. "Understanding and Mitigating Social Engineering Attacks to Elderly People: A Comprehensive Survey of Methods, Impacts, and Future Solutions." *2024 IEEE 15th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 2024.