3.5. Zakończenie

W tym rozdziale zaprezentowane zostały metody i rodzaje zagrożeń dla działalności biznesowej wynikających z ataków na systemy uczące się. Z przeprowadzonej analizy płyną dwa wnioski. Po pierwsze, ataki tego typu mogą w istotny sposób zaburzyć funkcjonowanie procesów biznesowych. Procesy biznesowe, wspierane sztuczną inteligencją, mogą zostać zmuszone do niepoprawnego działania. Ryzyko jest szczególnie wysokie w przypadku systemów, które mają wysoki poziom autonomii.

Po drugie, organizacje raczej nie uwzględniają specyfiki ataków na AI podczas zarządzania ryzykiem. Świadomość tych zagrożeń istnieje, jednak problemem jest brak narzędzi, które pomagałyby ograniczać ryzyko na etapie budowania i operacjonalizacji modeli AI (Kumar i in., 2020). W domenie sztucznej inteligencji istnieją jedynie zbiory dobrych praktyk i wskazówek, które mają na celu uchronić kod przed potencjalnymi lukami. Innych zabezpieczeń w zasadzie nie ma, choć specjaliści wskazują na konieczność uwzględniania sztucznie wygenerowanych „złośliwych" danych podczas uczenia modeli. Chodzi o to, aby modele były wyczulone na jak najwięcej tego typu przypadków (Dai i in., 2018).

Kontekst biznesowy ataków na systemy maszynowego uczenia się nie ogranicza się jednak do robotyzacji i automatyzacji procesów biznesowych. Obrana w tym rozdziale perspektywa ma charakter procesowy i pokazuje wiele aspektów funkcjonowania przedsiębiorstw, takich jak marketing, operacje, sprzedaż czy finanse. Dalsze rozważania związane z tego typu zagrożeniami powinny jednak objąć całość procesów biznesowych – od zakupów po sprzedaż.

Odmienny obszar potencjalnych zagrożeń stanowią szeroko pojęte zastosowania internetu rzeczy, szczególnie w dobie możliwości sieci 5G. Czujniki gromadzące dane na potrzeby inteligentnych samochodów, domów, miast czy inteligentnej produkcji, a także modele wykorzystujące dane z tych czujników też mogą stać się celem ataków przy wykorzystaniu antagonistycznych próbek danych.

BIBLIOGRAFIA

Andress, J. (2011). *The Basics of Information Security*, https://doi.org/10.1016/C2010-0-68336-2

Bhaumik, R., Williams, C., Mobasher, B. i Burke, R. (2006). *Securing collaborative filtering against malicious attacks through anomaly detection*. Center for Web Intelligence, DePaul University School of Computer Science, Telecommunication, and Information Systems Chicago, Illinois, USA.

Bielecki, W. (2001). *Informatyzacja zarządzania: wybrane zagadnienia*. Warszawa: PWE.

Cantos, M. (2019). Breaking the Bank: Weakness in Financial AI Applications, https://www.fireeye.com/blog/threat-research/2019/03/breaking-the-bank-weakness-in-financial-ai-applications.html (dostęp: 15.02.2020 r.).

Chawla, N.V., Bowyer, K.W., Hall, L.O. i Kegelmeyer, W.P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357, https://doi.org/10.1613/jair.953

Conrow, E. (2000). *Effective Risk Management: Some Keys to Success*. American Institute of Aeronautics and Astronautics.

Culkin, R., i Das, S.R. (2017). *Machine Learning in Finance: The Case of Deep Learning for Option Pricing Artificial Intelligence: A Reincarnation*. Santa Clara University.

Dabouei, A., Soleymani, S., Dawson, J., i Nasrabadi, N.M. (2019). Fast geometrically-perturbed adversarial faces. *Proceedings – 2019. IEEE Winter Conference on Applications of Computer Vision, WACV 2019*, 1979–1988, https://doi.org/10.1109/WACV.2019.00215

Dai, H., Li, H., Tian, T., Huang, X., Wang, L., Zhu, J. i Song, L. (2018). *Adversarial Attack on Graph Structured Data*, cyt. jako https://arxiv.org/abs/1806.02371

Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters.Com., https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G (dostęp: 15.02.2020 r.).

Davenport, T. i Harris, J.G. (2007). Inteligencja analityczna w biznesie: nowa nauka zwyciężania. *Archives*, http://books.google.com/books?id=n7Gp7Q84hcsC&pgis=1

Deldjoo, Y., Di Noia, T. i Merra, F.A. (2020). Adversarial Machine Learning in Recommender Systems. *AML-RecSys* (December), 869–872, https://doi.org/10.1145/3336191.3371877

Economist (2014). *Why the run on banks?* https://www.economist.com/eastern-approaches/2014/07/02/why-the-run-on-banks (dostęp: 11.02.2020 r.).

Fisher, M. (2013). Syrian hackers claim AP hack that tipped stock market by $136 billion. Is it terrorism? *Washongton Post*, https://www.washingtonpost.com/news/worldviews/wp/2013/04/23/syrian-hackers-claim-ap-hack-that-tipped-stock-market-by-136-billion-is-it-terrorism/ (dostęp: 15.02.2020 r.).

Goldblum, M., Schwarzschild, A., Cohen, N., Balch, T., Patel, A.B. i Goldstein, T. (2020). *Adversarial Attacks on Machine Learning Systems for High-Frequency Trading*, http://arxiv.org/abs/2002.09565

Goodfellow, I.J., Pouget-abadie, J., Mirza, M., Xu, B. i Warde-farley, D. (2014). *Generative Adversarial Nets*. Universite de Montreal.

Hacker News (2016). *Russian Hackers Manipulate Ruble-Dollar Exchange Rate with Malware*, https://thehackernews.com/2016/02/russian-exchange-hacked.html (dostęp: 10.02.2020 r.).

Hunt, E. (2016). Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter. *The Guardian*, https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter (dostęp: 11.02.2020 r.).

Koźmiński, A.K. (2004). *Zarządzanie w warunkach niepewności. Podręcznik dla zaawansowanych*. Warszawa: Wydawnictwo Naukowe PWN.

Kumar, R.S.S., Nyström, M., Lambert, J., Marshall, A., Goertzel, M., Comissoneru, A. i Xia, S. (2020). *Adversarial Machine Learning - Industry Perspectives*, http://arxiv.org/abs/2002.05646

Kwiatkowski, S. (2000). *Przedsiębiorczość intelektualna*. Warszawa: Wydawnictwo Naukowe PWN.

Lundberg, S.M. i Lee, S. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Mandel, R. (2017). *Optimizing Cyberdeterrence: A Comprehensive Strategy for Preventing Foreign Cyberattacks*. Georgetown University Press.

Minelli, M., Chambers, M. i Dhiraj, A. (2013). *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*. Wiley.

Moore, J., Hammerla, N. i Watkins, C. (2019). Explaining deep learning models with constrained adversarial examples. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, https://doi.org/10.1007/978-3-030-29908-8_4

Neekhara, P., Hussain, S., Jere, M., Koushanfar, F. i Mcauley, J. (2019). *Adversarial Deepfakes: Evaluating Vulnerability of Deepfake Detectors to Adversarial Examples*, http://doi.org/10.13140/RG.2.2.26227.48168

Niemimaa, M., Järveläinen, J., Heikkilä, M. i Heikkilä, J. (2019). Business continuity of business models: Evaluating the resilience of business models for contingencies. *International Journal of Information Management*, *49*, 208–216. https://doi.org/10.1016/j.ijinfomgt.2019.04.010

Niesen, T., Houy, C., Fettke, P. i Loos, P. (2016). Towards an Integrative Big Data Analysis Framework for Data-Driven Risk Management in Industry 4.0. *49th Hawaii International Conference on System Sciences (HICSS)*.

Papernot, N., McDaniel, P. i Goodfellow, I. (2017). *Practical Black-Box Attacks against Machine Learning*, cyt. jako arXiv:1602.02697

Perez, S. (2013). AP Twitter Hack Preceded By A Phishing Attempt. from https://techcrunch.com/2013/04/23/ap-twitter-hack-preceded-by-a-phishing-attempt-news-org-says/ (dostęp: 15.02.2020 r.).

Provost, F. i Fawcett, T. (2013). *Data Science for Business*. O'Reilly.

Ribeiro, M., Singh, S. i Guestri, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery.

Sadgrove, K. (2015). *The Complete Guide to Business Risk Management*. Routledge.

Schreyer, M., Borth, D. i Reimer, B. (2019). *Adversarial Learning of Deepfakes in Accounting*, cyt. jako arXiv:1910.03810

Sobczak, A. (2018). *RPA i AI w liczbach i cytatach*, https://blog.cionet.com/2018/04/18/rpa-i-ai-w-liczbach-i-cytatach/ (dostęp: 5.03.2020 r.).

Sobczak, A. (2020a). *101 pytań i odpowiedzi dotyczących robotyzacji biznesu*, https://robonomika.pl/101pytan/czym-jest-robotyzacja-procesow-biznesowych (dostęp: 12.04.2020 r.).

Sobczak, A. (2020b). *Zastosowanie sztucznej inteligencji w sektorze finansowym*, Forum Gospodarcze TIME website: https://robonomika.pl/zastosowanie-sztucznej-inteligencji-w-sektorze-finansowym-moja-prezentacja-z-forum-gospodarczego (dostęp: 12.03.2020 r.).

Surma, J. (2009). *Business Intelligence*. Warszawa: Wydawnictwo Naukowe PWN.

Surma, J. (2017). *Cyfryzacja życia w erze Big Data*. Warszawa: Wydawnictwo Naukowe PWN.

Suwajanakorn, S., Seitz, S.M. i Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: Learning Lip Sync from Audio. *ACM Transactions on Graphics*, *36*(4).

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I. i Fergus, R. (2014). Intriguing properties of neural networks. *International Conference on Learning Representations*.

Tupa, J., Simota, J. i Steiner, F. (2017). Aspects of risk management implementation for Industry 4 0. *Procedia Manufacturing*, *11*, 1223–1230. https://doi.org/10.1016/j.promfg.2017.07.248

Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, *9*(11).

Zawiła-Niedźwiecki, J. (2013). *Zarządzanie ryzykiem operacyjnym w zapewnianiu ciągłości działania organizacji*. Kraków–Warszawa: edu-Libri.

Zhou, W., Wen, J., Qu, Q., Zeng, J. i Cheng, T. (2018). Shilling attack detection for recommender systems based on credibility of group users and rating time series. *PLoS ONE*, *May*, 1–17.