update date: 23.06.2022

#	Research topics offered by IDEAS NCBR	Abstract	Group/team leader in IDEAS NCBR	E-mail
1	**More secure decentralized finance**	One of the main problems of cryptocurrencies is their highly fluctuating exchange rates with standard currencies. Decentralized finance (often abbreviated as DeFi) refers to blockchain solutions that address this problem. The most notable examples are the stable coins, where the internal currencies of a given blockchain are exchangeable with fiat money at a fixed rate. Several exciting proposals exist for constructing such coins, most coming from blockchain startups and lacking complete formal security analysis. This student will work on improving these protocols and understanding their security properties. DeFi protocols can also be used to replace financial institutions such as the stock market. One of the problems with this approach is the so-called front-running attacks. Unfortunately, most blockchain solutions allow such powerful participants to publish their transactions on the blockchain before the original transactions appear there. This can lead to considerable financial losses for such honest participants. This student will work on addressing this problem.	Stefan Dziembowski	stefan.dziembowski@ideas-ncbr.pl
2	**Proofs-of-Space in practice**	The most popular blockchain platforms use consensus based on the so-called Proofs-of-Work, where the participants are incentivized to constantly solve many computational puzzles (this process is also called mining). This leads to massive electricity consumption. Several alternatives to Bitcoin mining have been proposed in the past. Stefan Dziembowski (who leads this research at IDEAS) is one of the authors of another approach to this problem, called the Proofs-of-Space. In this solution, the computational puzzles are replaced with proofs that a given party contributed some disk space to the system. Several ongoing blockchain projects are based on these ideas. This student will work on improvements to these protocols.	Stefan Dziembowski	stefan.dziembowski@ideas-ncbr.pl
3	**Real-life side-channel security of blockchain wallets**	Another critical weakness in the vision of decentralizing internet services is that interacting with blockchains is more complicated than in the case of centralized solutions. Moreover, the decentralization makes it impossible to revert the transactions that were posted by mistake or as a result of an attack. Due to this, the users often rely on the so-called hardware wallets, which are dedicated devices protected against cyber-attacks. This student will work on analyzing the security of the existing hardware wallets. In particular, we will be interested in their side-channel security, i.e., security against attacks based on information such as power consumption or electromagnetic radiation.	Stefan Dziembowski	stefan.dziembowski@ideas-ncbr.pl
4	**Practice-oriented verification of blockchain protocols**	One of the main problems in the blockchain space is that decentralized solutions are typically more complex and error-prone than centralized ones. In particular, errors in smart contracts can lead to considerable financial losses. Furthermore, some blockchain algorithms in the past had serious mistakes that could be used to steal large amounts of money. This student will work on addressing these problems using tools from formal methods, in particular, proof assistants and provers such as Coq, Easycrypt, Why3, and others. Theoretical aspects of this work are one of the topics of the ongoing ERC grant of Stefan Dziembowski. This student will work on more technical aspects, especially making this approach usable in real life by blockchain developers.	Stefan Dziembowski	stefan.dziembowski@ideas-ncbr.pl
5	**Explainable Algorithmic Tools**	In this research project we aim to propose tools that would provide explanations for the different basic optimization problems, e.g., assignment problem, shortest paths, minimum cuts, or basic graphical neural networks. This research is motivated by the fact that even when faced with problems that can be solved exactly, we still would like to understand why this solutions was computed.	Piotr Sankowski	piotr.sankowski@ideas-ncbr.pl
6	**Learned Data-Structures**	ML tools enter as interior components into basic data structures or state-of-the-art approximation algorithms resulting in solutions that have better practical properties, e.g., indices. These new hybrid constructions are called learned data-structures. As the work on these ideas has just started we miss the right framework and tools for implementing state-of-art solutions and thus the research on new tools and models is hampered. This research aims to continue research on this problem and create new algorithms and data structures together with their implementations. This could prove tools to bridge the gap between theory and practice in algorithms and show that new theoretical advances can have practical implications.	Piotr Sankowski	piotr.sankowski@ideas-ncbr.pl
7	**Algorithmic Tools for Data Science and ML**	Although, different parallel computation models have be studied for years already. A new model that describes real-world systems has been proposed recently - the Massively Parallel Computation (MPC) frameworks includes systes such as MapReduce, Hadoop, Spark, or Flume. It comes with a completely new possibilities as well as requirements. MPC computations are executed in synchronous rounds, but implementing these rounds on real-world systems takes considerable time. One round takes orders of magnitude longer than on classical Cray type system. Thus we would like to solve problems, in particular graph problems, in as few rounds as possible. With this challenge in mind, this project aims to design methods to break barriers that were impossible to overcome using classical techniques and models. More specifically, we are going to work on new algorithmic tools that would improve efficiency of both parallel and non-parallel algorithms used in data science.	Piotr Sankowski	piotr.sankowski@ideas-ncbr.pl
8	**Deep NLP Models for Polish Language**	In recent years we observe a huge progress in development of deep NLP models. In many applications these models can effectively compete with humans, and there usage is growths. However, the main works on these models are limited to mayor languages, and recent developments are not directly available for Polish language. The aim of this project is twofold: develop cutting edge NLP models for Polish language; use the experience gained this was to extend and improve models for other languages.	Piotr Sankowski	piotr.sankowski@ideas-ncbr.pl
9	**Universal and Multi-modal Neural Networks**	In this project we aim to work on multi-purpose and multi-modal neural networks. The tasks we aim to cope with will different problems where we aim to integrate different kind of information and aim to deliver joint representation that would allow for example: translate text to images and vice-versa for general and medical usage; transform natural language to animations, or approach no-code programming challenges.	Piotr Sankowski	piotr.sankowski@ideas-ncbr.pl

10	**Adversarial Social Network Analysis**	How can individuals and communities protect their privacy against social network analysis techniques, algorithms and other tools? How do criminals or terrorists organizations evade detection by such tools? Under which conditions can these tools be made strategy proof? These fundamental questions have recently attracted increasing attention in the literature as a new paradigm for social network analysis, whereby the strategic behaviour of network actors is explicitly modeled. Addressing this research challenge has various implications. For instance, it may allow two individuals to keep their relationship secret or private. It may also allow members of an activist group to conceal their membership, or even conceal the existence of their group from authoritarian regimes. Furthermore, it may assist security agencies and counter terrorism units in understanding the strategies that covert organizations use to escape detection, and give rise to new strategy-proof countermeasures.	Tomasz Michalak	tomasz.michalak@ideas-ncbr.pl
11	**Game-Theoretic Aspects of Blockchain Technologies**	The blockchain is a game changing technology and as such it has attracted enormous interest from both academia and industry. While there are countless potential application of blockchain, almost all of them share a common feature: the parties that use it are assumed to be, in principle, self-interested utility maximizing individuals. Given this, many aspects related to the blockchain technology should be analysed using the aparatus of game theory. These include such issues like: selfish mining, majority attacks and Denial of Service attacks, computational power allocation, reward allocation, and pool selection, and energy trading. While the literature that analyses game-theoretic aspects of blockchain is growing, there are many interesting open questions that have not yet been answered in a satisfactory way. For instance: how to design rules that lead to the development of payment channel networks that are secure, reliable and efficient.	Tomasz Michalak	tomasz.michalak@ideas-ncbr.pl
12	**Stackelberg Games for Security**	Various global threats, including terrorism, drug and human trafficking have led the researchers to look for more efficient way to utilize security resources. An interesting approach that delivered surprisingly good results is based on game theory. Instead of deploying security forces by expert decisions, an optimal solution of an appropriately constructed security game is used. This approach – based on Stackelberg Security Games – has proven to deliver substantially better results. It has been applied in numerous places in the USA (Los Angeles International Airport, the US Federal Air Marshals Service, Boston harbour) and all over the world (to endangered species in natural parks). In this research line, we would like to deploy such solutions in Poland. To this end, given the increasing number of threats and their changing nature, there is a need to extend the current theory of Stackelberg Security Games as well as the algorithms to compute them. We plan to deploy potential results in various critical places in Poland.	Tomasz Michalak	tomasz.michalak@ideas-ncbr.pl
13	**Explainability of Machine Learning Models Based on Game Theory**	One of the key research challenges regarding machine learning models is their explainability. When high-value decisions are taken, e.g., in medical diagnostic, understanding why a model made a specific prediction is often as important as the prediction's accuracy. Thus we need to develop methods to interpret the model's results in a transparent way so that humans are willing to follow model recommendations. As a result, recently, there has been a growing interest in the feature attribution problem, where, given some specific input x of features, one would like to attribute the model's prediction f(x) to the individual features. One of the most popular approaches to interpreting model predictions uses methods originating from cooperative game theory that are called solution concepts or values. They measure the importance of each player in, or contribution to, a coalitional game. While there exist many ways in which the importance of each player can be evaluated, some solution concepts are considered more fundamental than others due to underlying axiom systems that uniquely determine them. One important game-theoretic solution concept that attracted a lot of attention in the context of explainability is the Shapley value popularized by the SHAP library in python. However, the Shapley value is not the only solution concept that has been advocated for interpreting model predictions. Some papers suggest using other values. Each value has its own unique characteristics and should be used for specific applications and types of machine learning models. It is then a very pressing to do a thorough study of game-theoretic solution concepts for explainability of machine learning models. The challenge involves: (a) the study of theoretical aspects of applying various game-theoretic solution concepts; (b) the computational analysis---given the inherent computational challenges related to game-theoretic solution concepts it is paramount to look for tractable approaches to the problem; and (c) experimental analysis in chosen applications; and (d) possible implementation as a software library.	Tomasz Michalak	tomasz.michalak@ideas-ncbr.pl
14	**Continual learning of neural networks**	Despite the recent successes in the fields of image, text, and sound processing, based on neural networks, adapting the models to changing data conditions still poses a significant challenge. Continual learning is a discipline that deals with the problem of changing the characteristics of the data used to train a model over time. The most important challenge is catastrophic forgetting, which causes the model learned sequentially on two datasets to lose its accuracy on the former with training on the latter. The project will develop methods for training deep neural networks that can address the problem of forgetfulness and create new application possibilities for continual learning.	Tomasz Trzciński	phd@ideas-ncbr.pl
15	**Learning data representations for computer vision and machine learning**	Various data representations are crucial for solving multiple real-life applications, including autonomous driving, robot manipulations and language processing. In this project, we plan to develop novel methods for learning data representations leveraging neural network architectures. We will focus specifically on visual and multimodal representations and investigate methods using supervised and unsupervised (e.g. generative) models to that end.	Tomasz Trzciński	phd@ideas-ncbr.pl
16	**Zero-waste machine learning**	The computations run by contemporary machine learning models to process the increasing amount of data come at an enormous price of long processing time, high energy consumption and large carbon footprint generated by the computational infrastructure. Moreover, neural networks become increasingly complex, which leads to high monetary costs of their training and hinders the accessibility of research to less privileged communities. Existing approaches to reduce this burden are either focused on constraining the optimization with a limited budget of computational resources or they attempt to compress models. In this project, we plan to look holistically at the efficiency of machine learning models and draw inspiration to address their main challenges from the green sustainable economy principles. Instead of limiting training of machine learning models, we want to ask a different question: how can we make the best out of the information, resources and computations that we already have access to? Instead of constraining the amount of computations or memory used by the models, we focus on reusing what is available to them: computations done in the previous processing steps, partial information accessible at run-time or knowledge gained by the model during previous training sessions in continually learned models.	Tomasz Trzciński	phd@ideas-ncbr.pl

17	**Partial information in self-supervised learning**	Self-Supervised Learning (SSL) was introduced as a remedy for massive amounts of labeled data required by supervised approaches to building intelligent generalized models. It exploits the freely available data to generate supervisory signals which act as labels. For this purpose, in the case of image classification, SSL uses different image distortions, also referred to as augmentations. While self-supervised approaches provide on par or superior results to their fully supervised competitors, they are computationally demanding, requiring large batches or momentum encoders. This project aims to leverage partial information into self-supervised strategies to increase their efficiency and reduce computational costs. Partial information assumes that a set of labels corresponding to a given image is known during inference, and it can be used to improve the performance of the model. This corresponds to a real-life application, where, for instance, we know that the image was captured in a forest or in a cave. The proposed research topic will leverage partial information into SSL, among others, by developing augmentation methods that use contextual information as a distortion source and utilize it as supervision in self-supervised learning.	Tomasz Trzciński	phd@ideas-ncbr.pl
18	**Partial information in attention-based models**	Ever since the transformer was introduced in 2017, there has been a huge success in the field of Natural Language Processing (NLP). The main reason for the effectiveness of the transformer is its ability to handle long-term dependencies compared to RNNs and LSTMs. After its success in NLP, there have been various approaches to its usage for Computer Vision tasks. However, while transformers provide state-of-the-art results, they require large-scale training to trump an inductive bias. This project aims to leverage partial information into attention-based models to increase their efficiency and reduce computational costs. Partial information assumes that a set of labels corresponding to a given image is known during inference, and it can be used to improve the performance of the model. This corresponds to a real-life application, where, for instance, we know that the image was captured in a forest or in a cave. The proposed research topic will leverage partial information into attention-based models, among others, by incorporating partial evidence to model sparsity in attention layers.	Tomasz Trzciński	phd@ideas-ncbr.pl
19	**Is self-learning really enough? Unsupervised representation learning.**	In this project, we will study the problem of building efficient representations for downstream tasks in a continual learning scenario. Recently, novel self-supervised approaches showed promising results when they are properly regularized. We will investigate how internal network representation can be prepared for best re-use in the downstream tasks when trained continuously without supervision. Such an approach can be applied later to many downstream tasks in a cost-efficient way, i.e. with only a simple fine-tuning of a small and task-dedicated part of the model.	Tomasz Trzciński	phd@ideas-ncbr.pl
20	**Efficient experience usage for life-long learning. Are exemplars all you need?**	Storing exemplars directly in an additional memory buffer is the most common way to get acceptable performance in continual learning tasks. This allows you to easily learn cross-task features. Exemplar-based methods for class incremental learning or experience replay methods for online continual learning are focused on the efficient use of a given memory buffer by appropriate selection and retention of exemplars. Different methods directly optimize stored exemplars or use given memory to store models that allow generating samples or features -- the so-called pseudo rehearsal. The research question we ask in this project is the following: Is there a way to store previous knowledge more efficiently? Can we prompt saved representation in memory better, i.e. learn to prompt or query it?	Tomasz Trzciński	phd@ideas-ncbr.pl
21	**Representation alignment What should not be regularized in continual learning?**	Regularization-based methods are one of the easiest to apply and most common techniques for incremental learning, where we cannot store exemplars. There are two main types of regularization techniques: based on the weights like EWC or on the network outputs, such as LwF. Both try to alleviate the problem of catastrophic forgetting by keeping the network regularized -- enforcing the current network to remain similar to the old model and be able to solve similar tasks to the old ones by that. This increases the stability of the model, and can therefore hurt plasticity. In this research topic, we attempt to answer the following research questions: Is there a good trade-off for that? Maybe some aspect of the network should not be regularized at all, or regularized in a completely different way?	Tomasz Trzciński	phd@ideas-ncbr.pl
22	**Cooperate to learn continually**	Most class incremental learning methods assume one network, one backbone encoder to solve all the tasks, previously seen and the new ones. The signal goes through all the networks to solve any task. In living organisms, this is not exactly the case. Sensory information goes through different compartments that focus on various aspects of input signals. In addition, they are coordinated by a more global signal, e.g. gated by dopamine. In this line of research, we would like to focus on the cooperation of many learners - usually smaller, more energy-efficient, and weaker in comparison to the one-big model. Continual learning of them needs additional coordination.	Tomasz Trzciński	phd@ideas-ncbr.pl
23	**Continual Federated Learning**	In Federated Learning (FL) we have a central server node and many peers - clients that learn on their own data. We exchange only the model gradients to and from the server. Clients do not share their data or any information that can break privacy. Usually, a differential privacy model is applied to enforce that. This is an attractive way to train models in many domains, e.g. healthcare, advertisements, and mobile applications, just to name a few. Most of the use cases are based on the static data, split for clients, and then the FL training process proceeds. Neither tasks nor data are changing along the way. New concepts are not emerging at the client's level. Simply, they will be averaged (FedAvg) and lost. Most of the methods do not consider learning anything and how to integrate and propagate this knowledge from the server to other clients. In this work, we address the problem of incremental learning on clients' devices, usually edge devices, low-energy, and memorable ones. Learning new concepts in such an environment is challenging and not well explored so far.	Tomasz Trzciński	phd@ideas-ncbr.pl