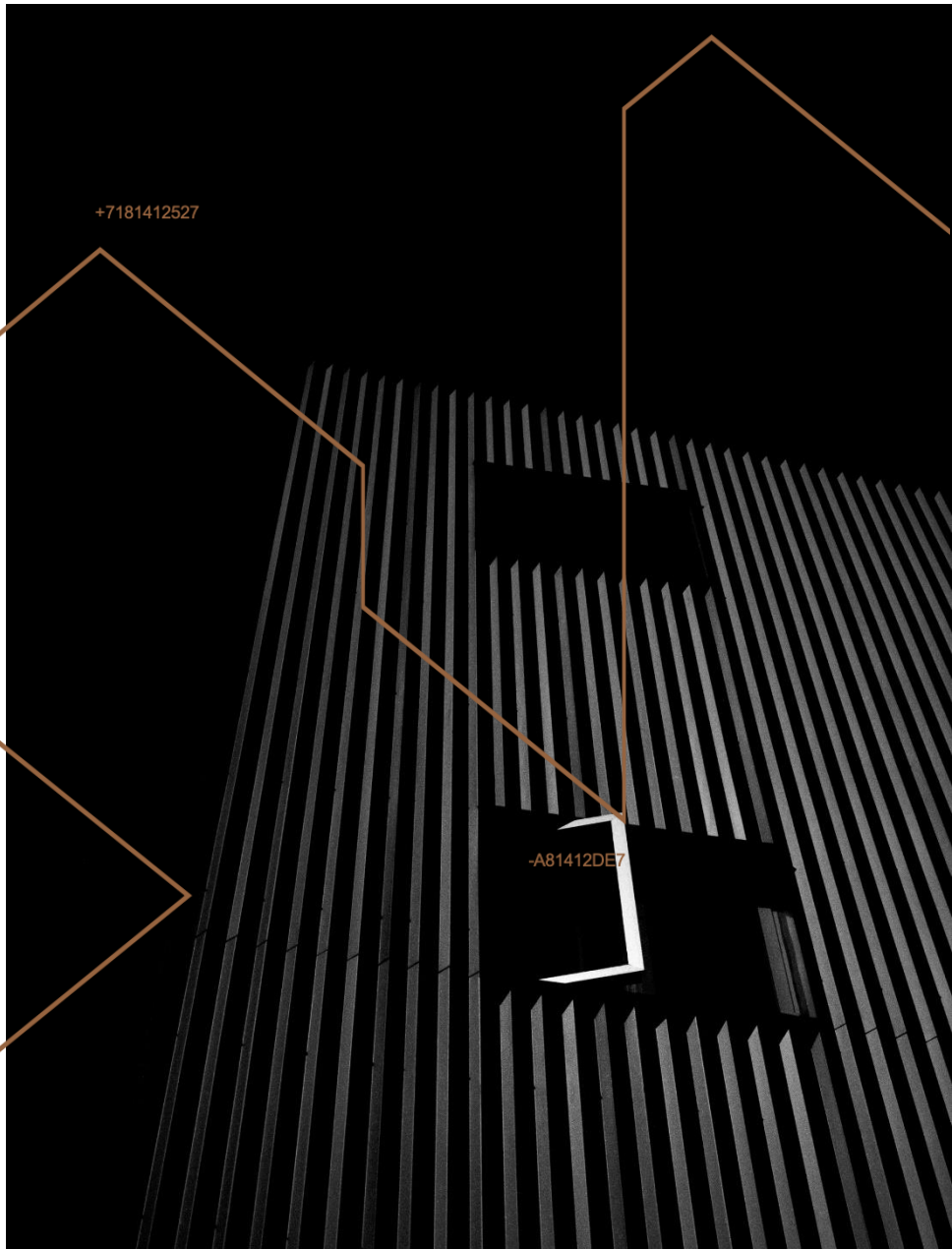
guardtime

guardtime.com

Convergence of Blockchain and Artificial Intelligence

by Guardtime Research

Whitepaper
July 2019



INTRODUCTION

There are several characteristics of blockchain (or more generally, distributed ledger) technology that are of potential interest for artificial intelligence applications. The decentralized nature of blockchains allows for convenient sharing of AI training data, training processes and of pretrained AI models. While in the past, traditional databases and repositories have been used for sharing data, the trustless nature of blockchains, along with the transparency, immutability and auditability provided via cryptographic provenance allow for secure sharing, with improved trust on the AI models and the data they work with. Blockchains can also serve as a global registry for AI data and models, with appropriate permissions and access controls enforced.

This paper provides a brief background on blockchains and AI applications, how they can benefit from each other and also debunks some of the myths on the synergy between the two.

BACKGROUND

AI can be divided into two broad categories: general artificial intelligence (i.e., artificial human-like intelligence) and specific applications of artificial intelligence (building applications that possess some intelligence for solving problems in a narrow field). General artificial intelligence is out of reach of current science and technology and rather belongs to the field of science fiction and philosophy than to an area of practical science and applications. Based on present common knowledge, this situation will remain so for the foreseeable future. Hence, we will not consider the first category here and will focus on the second, more practical category. The applications from the second category are being built using techniques from the following areas: knowledge handling, search, learning and decision making, and problem solving and planning.

It is important to understand that from a technical perspective there is no fundamental difference or distinction between AI programs and ordinary non-AI programs. In general, both types of programs consume some input data, do some processing, and produce an output. The processing is done with algorithms of varying complexity and sophistication which are implemented in some programming language. This applies to both training of AI models (training data in, model out) and using them (actual data in, decisions out).

Although in a lot of contexts, machine learning is used synonymously with artificial intelligence, it is actually just a family of techniques (e.g., statistics) for computing a compact representation of information contained in a larger data set. Here, we will specifically consider the artificial intelligence topic in the context of machine learning.

We consider AI algorithms (or, also called models) that are automatically derived in the process of statistical data analysis (or, simply speaking, from data) rather than by hypothetically driven development, like most software is being developed nowadays. Machine learning deals with the methodology of generation of software black boxes in the process of learning from application case-specific sample (or training) data sets. The purpose of these black boxes may be to detect familiar patterns in input data (detecting and recognizing objects in images and video streams, speech recognition, detecting cyberattacks in IT infrastructure, etc.), detecting and classifying regularities in generally unstructured data (market segmentation, fraud detection, identifying trending topics from social media, etc.), generating predictions from observed data (weather prediction, stock market prediction, personal behavioral prediction for efficient marketing, etc.), decision making (autopilots, robots, investment advisor bots, etc.), and others. Basically, in any use case dealing with large amounts of complex data (Big Data) where conventional algorithm development and software building methods are impractical, machine learning approaches may be in demand.

The process of training a machine learning model is quite complex, iterative and highly computationally intense. The quality of a resulting model (i.e., how close its outputs fit the expected results) depends not only on the chosen training data sets, but also on an approach that was used to train the model (how many iterations, how training data set was chosen, how it was split between iterations, how model quality and fitness was measured, etc.). It also depends on the particular version and type of software "machine learning toolkit" that was used to derive the model. Since building a machine learning model from scratch for each particular use case may be prohibitively difficult, it is more practical to take already existing pretrained models for broadly defined domains and train them further for a particular use case (so-called *knowledge transfer*). There exist a number of freely available libraries with such generic pretrained models (for instance [7,8,9,10,11,12]).

Consider, for instance *Common Objects in COntext - Single Shot Detection (*COCO SSD) pretrained models that can be customized for image object detection applications [7,8,9]. COCO SSD solves a core computer vision task- build a machine learning model that would be able to recognize multiple object recognition and localization on a single image. It is extremely challenging and a computationally intensive task to build such kind of a model. COCO SSD [9] is built on top of Google's Tensorflow Object Detection API model [7] and trained to detect objects from COCO dataset [8]. The API, while based on TensorFlow, is a machine learning framework used to construct, train and deploy object detection models. COCO contains large-scale object detection, segmentation, and captioning dataset. Out-of-the-box, COCO SSD model can detect over 80 categories of real-world objects in images. Since COCO SSD is already well trained to recognize and locate different types of objects or images under different conditions and within different contexts, it is relatively easy to tune (train further) COCO SSD to recognize/locate narrower type or completely new type of use-case specific objects.

Thus, the end product model will depend on training data, training process, software environment and more generic initial pretrained models that were customized to obtain this model.

One of the greatest challenges at present in the field of machine learning is siloed data. In modern world, competing organizations are reluctant to share with their competitors any piece of information or technology that would anyhow benefit their peers. Hence the training data for AI's that are being built for similar tasks in different organizations is being kept in secret and isolated from others. Meanwhile, an AI model performs much better if it is being trained on a larger highly diverse data set. Hence, there exists a need for an ecosystem (related to federated learning) that would facilitate cross-organizational data set and AI model sharing and reuse, while not breaking NDAs, not revealing corporate secrets, and assuring clean uncompromised shared data sets and models.

In the context of the blockchain and artificial intelligence discussion, other, non-technical aspects of the AI concept seem to dominate. The main concern appears to be making sure computers will not make bad or dangerous decisions when people delegate decision making to software (AI). In specific applications, there may exist heuristics to check if an automated decision is good and safe before actually executing it. Unfortunately, for non-trivial decisions (and the majority of AI applications will be non-trivial), this is impossible to do (otherwise the AI software would incorporate the rules for checking its output). On the other hand, validators (miners) in blockchain networks just enforce (often rather trivial) policies and rules of the network by executing a validation algorithm. Along with using blockchains, software development best practices and state of the art quality assurance techniques can and should be applied to AI application development (for example [2,3]).

The term blockchain is used in non-technical discussions in various meanings. Ignoring technical details, one could think of a blockchain as a distributed system consisting of a shared append-only journal of transactions. The implementation guarantees that all participants have a consistent view, new entries added to the journal conform to the validation rules and the history is tamper evident. Thus, the participants can trust the system to operate according to its specification.

Blockchain-based ledgers are said to be single sources of truth. "Truth" cannot refer to factual correctness of data, because a program has no way to validate the factual correctness of input data. It cannot refer to whether data is current and up to date either, because, for example, external systems may delay the registration of up to date information in the blockchain system. "Truth" can only mean the data is correct in terms of having its integrity preserved through its life and all parties' view of the data being consistent.

BLOCKCHAIN FOR AI

There are some advantages of using blockchain in conjunction with AI applications. Trustworthiness of data is key to any AI application. Trustworthy AI [1] has several components to it, but blockchain can aid in certain aspects like governance, transparency via traceability, and accountability via auditability. A crucial component of trustworthy AI is robustness. It should be noted that blockchains cannot prove the correctness of the data, only the latest possible modification time and the absence of unauthorized changes can be proved with cryptographic evidence produced by a blockchain.

Common Myths:

+ ***Autonomous systems and machines based on smart contracts can learn and adapt to changes over time and make trusted and accurate decisions that are verified and validated by all mining nodes of the blockchain. Such decisions cannot be refuted, and can be traced, tracked and verified by all participating entities.***

In blockchain networks, mining nodes (validators) are able to verify that all transactions follow the protocol of the network and that the instructions defined in smart contracts were executed according to the semantics of the smart contract language and execution environment. These relatively straightforward and fixed rules are coded into the client software used by miners. Unfortunately, miners have no means for verifying or validating the correctness of the instructions of smart contracts relative to intentions of users. Therefore, a programming mistake or incorrect logic implemented in a smart contract, while being in compliance with the blockchain protocol and executed according to the semantics of the programming language, will lead to unexpected outcomes. In other words, a blockchain that supports general purpose programmable smart contracts will support and execute both “correct” and “incorrect” smart contracts without the ability to distinguish the two categories. The inability to refute incorrect outcomes, stemming from the immutability of the blockchain, can lead to serious negative consequences as manifested by several cases of theft of cryptocurrency. Moreover, miners do not have the capability of assessing the correctness of input data used by smart contracts, which is another potential source of unintended outcomes. Execution of smart contracts (similar to any computer programs) can indeed be traced instruction by instruction, however, decisions made by AI applications are likely to consist of a large number of steps, and the structure of the models obtained by machine learning are not optimized for human comprehension, and thus vetting of contracts before their deployment to the blockchain is not realistic either. In summary, even leaving aside the observations that computations on blockchains tend to be slow and inefficient, and storage is relatively expensive, in the light of previous discussion, blockchains are unlikely to be a reasonable platform for autonomous, adaptable smart contract-based AI applications.

+ ***Efficient decentralized training/learning models.***

It has been speculated that blockchain may enable an efficient way of running machine learning algorithms on isolated data silos owned by different parties not willing to share the data, and then composing resulting models into one more powerful model that can be shared (federated learning). While blockchain can be used for proving the integrity of the training data, training algorithms and resulting models, the model composition problem is completely unrelated to blockchain technology. It has been shown that under some assumptions (e.g., trusted hardware) federated learning can be implemented using a blockchain instead of a central server [6], but the system becomes less efficient and the security benefits are unclear.

+ ***Speed up hashing calculations.***

In the context of using AI to improve blockchains, there is a wrong assumption that slow hashing is the bottleneck limiting the transaction throughput of public blockchain networks (e.g., Ethereum, Bitcoin). However, the slowness of the Proof-of-Work hash calculation is by design intended to slow down the consensus process for ensuring proper data propagation through the network. In reality, even if faster implementations of the hash function used in these blockchain networks were invented, the consensus algorithms would adjust the difficulty level of Proof-of-Work puzzles to require more hash calculations in order to keep the average block time in a fixed range. Hence, using AI for optimizing hash function implementations will have no actual benefit for the scalability of blockchain networks or improving energy consumption.

+ ***AI and encryption work well together.***

AI and encryption are rather orthogonal technologies. While technically this statement can be considered correct, it contains no useful insights. Also, encrypted data (be it in the AI conext or otherwise) always had additional overhead compared to the plaintext.

+ ***Blockchain protects data privacy.***

Privacy isn't something that comes by default when data is stored on the blockchain, because by the nature of a blockchain system all participants will have an exact copy of the whole shared database. Privacy has to be carefully layered onto a blockchain based system. Various factors need to be considered - who has read/write access to the data, who can see the transactions (private channels), who has the permissions to execute smart contracts etc. Privacy in this context refers to privacy of the participants (anonymity) and data privacy.

+ ***Blockchain makes systems more efficient and scalable.***

A system can be made more efficient and scalable by adding a component, only when this involves some architectural changes towards a better design, otherwise the additional component will have zero or negative effect because of the overhead it inevitably introduces. Blockchain-based systems usually have high overhead because of additional communication needed for establishing consensus. Moreover, in order to be able to verify the integrity of the history of the journal of transactions, historic data has to be kept available, requiring potentially large amounts of storage space. Generating and validation cryptographic proofs can also require significant computational resources. In order to justify the use of a blockchain, the tradeoff between loss of efficiency and scalability and other benefits the blockchain brings has to be assessed. A notable exception here is the KSI

blockchain based timestamping that can be implemented with almost zero overhead. In general, however, the claim that blockchain makes systems more efficient and scalable is false.

+ ***Blockchain makes AI secure.***

Security in general is not composable, i.e., adding a secure component or layer to a (potentially insecure) system will not make the system secure as a whole. Often, the opposite is true - adding a component opens additional attack vectors and introduces unexpected side-effects affecting negatively data confidentiality, integrity or availability. Therefore, the claim that adding blockchain to AI makes it secure is false.

+ ***Blockchain helps prevent unintentional bias in the training data.***

Preventing unintentional bias in learning data is a challenging task. Bias can be detected by an entity able to compare the statistical measures of the training data to the "real world". Blockchains have a very limited interface to the external world, often implemented as "oracles" which are trusted sources for information about the external world. Using a blockchain for computing statistical measures over large data sets for detecting bias is a rather inefficient and complicated way (probably even infeasible for non-trivial cases) to achieve this goal and without any clear benefits.

GUARDTIME'S KSI BLOCKCHAIN FOR ASSURED AI

Guardtime's KSI blockchain is designed to provide massively scalable digital signature-based authentication for electronic data, machines and humans. The unique "keyless" signature or tag can be stored with the data or separately and aids in verification of the time of creation, identity of creator, and integrity of the data. Unlike traditional approaches that depend on asymmetric key cryptography, KSI uses only hash-function cryptography, allowing the verification to rely only on the security of hash-functions.

Guardtime's KSI blockchain, in conjunction with supporting middleware, offers a cryptographically sealed provenance thereby rendering all critical assets and the evidence of interactions between users and these assets immutable. KSI signatures are verified in real-time, with notifications sent, should data integrity be compromised and / or unauthorized access occur. KSI-based detection and attribution is widely witnessed owing to the calendar publication and hence it is possible for malicious activity to be *provably* detected and communicated. The ability to rapidly extract chain of custody evidence without exposing critical information from the digital assets under KSI-protection makes same-day forensic responses now possible.

Specifically, in the context of AI, KSI can offer:

/ RESILIENCE TO ATTACK

AI systems, like all software systems, should be protected against attacks like poisoning of training data and pretrained models by outside entities and corruption of the AI applications leading to erroneous decision making with unintended results.

/ ACCURACY

While accuracy of AI predictions cannot be guaranteed by the mere use of blockchains, it can be enhanced by ensuring the integrity of the data and pretrained models. Malicious actors in the system can be deterred via cryptographic provenance of all actions/decision making processes.

/ QUALITY AND INTEGRITY OF DATA

Quality of data is crucial. It is imperative that data be 'clean' before getting registered to the blockchain. While the quality of data cannot be assured by the use of a blockchain, integrity protection on the data can help identify corruption. AI decisions based on integrity protected but incorrect (e.g., biased) data are very likely also incorrect. Corrupt data will change the behavior of the AI systems and hence integrity protection of both the processes and data, along with the protection of the software supply chain for the AI systems (be it developed in-house or outsourced) is key. Blockchain-based solutions can be used for proving the integrity (the absence of unauthorized changes) of data that AI-powered systems rely on.

/ ACCESS CONTROL

Blockchain-based distributed ledgers can be used for storing records defining access rights. When information about access rights are stored in a blockchain, external systems actually enforcing the rights have to consult the blockchain for up to date information. The advantages of using a blockchain instead of a conventional database for this use case are unclear. When the training data itself is stored in the blockchain (which is relatively inefficient and expensive), every node in the blockchain network will have a copy of the data and can technically use the data without any restrictions. Even if the data is stored in the blockchain encrypted, it is difficult to enforce access control, because decryption keys may leak, or encryption scheme broken, and the data cannot be removed from the blockchain. Therefore, blockchain could be used as a shared database of records of access rights, but blockchain itself cannot enforce access control rules. KSI blockchain, while not providing the functionality of a ledger, may be used for generating integrity proofs on the access control lists.

/ TRANSPARENCY

Traceability, auditability and explainability of the AI decisions can be facilitated by utilizing the provenance on the data set changes/access, training processes, used software and initial pretrained models. According to the Ethics Guidelines for Trustworthy AI report [1], technical explainability requires that the decisions made by an AI system can be understood and traced by human beings. Blockchain based audit trail of input data and decision-making processes enables traceability and hence explainability to some degree. While blockchain cannot prevent an AI-based system from failing in unexpected ways [5], blockchain-based evidence can be used for after the fact analysis. Blockchain provides the ability for users to have visibility into who accessed their data. It also provides a cryptographically sealed linkage, tamper resistant for the actions taken on the data and the data itself.

/ ACCOUNTABILITY

Data stored on the blockchain lends itself to audability via immutability.

/ COMPLIANCE

AI rules (white lists/blacklists) can be registered to the blockchain. This ensures that the rule set is immutable, thereby guaranteeing that the AI system doesn't have behavior transgressions.

/ PROCESS AUDIT

Blockchain can be used to record AI-powered decisions (or decisions made by any other means) in order to provide auditability. Due to the relative inefficiency of blockchains, using a blockchain directly for audit logs is likely impractical. A better approach could be the use of conventional logging solutions for audit logs, which also offers confidentiality and then applying KSI to prove the integrity of audit logs.

/ DATA/MODEL SHARING

Blockchain-based ledger can serve as a shared and trusted repository for anchoring hash values of large data sets, training processes, software libraries and pretrained models used for machine learning or other AI applications. The data itself can be stored employing common and more efficient storage technologies. Storing large data sets in blockchain is inefficient and may conflict with regulations (e.g., GDPR) even when the data is encrypted. Hash values stored in blockchain can be used to validate the integrity of the data obtained from some external storage system. KSI blockchain, while not providing the functionality of a ledger, may be used for generating integrity proofs.

BLOCKCHAIN, DIGITAL TWINS AND AI

Oftentimes, AI and machine learning are used to analyze the model of operations represented by a digital twin. With digital twins, for every physical asset, there would be a virtual copy of it (a digital duplicate that is not intended to replace a physical object), possibly running in the cloud. Data is fed to the virtual equivalent, any lessons learnt are then applied to its physical twin to aid in optimizations, predictive modeling, etc. Digital twins consume historical context and data to review present conditions and to apply machine learning/AI to predict the future. Blockchains could provide the immutable, trusted and distributed data-synchronization bus needed to tie together different elements of the digital twin environment. In a digital twin environment, blockchains can provide proof of provenance to track and trace the supply chain, and also be used to provide machine integrity.

In order to maximize their usefulness, digital twins need to be powered by high-performing databases that can pull together and process many data sets in real-time. A KSI type blockchain can protect the integrity of the database, the logs and also the integrity of the data shared between digital twins, used for predictive modeling.

BLOCKCHAIN ENABLED AI IN VARIOUS SECTORS

HEALTHCARE

While Electronic Healthcare Records (EHR) are a goldmine of training data for data-hungry AI applications in healthcare, it is imperative to provide guarantees on the integrity and quality of the data that the predictive algorithms will work on. Blockchains in conjunction with AI can aid in personalized precision healthcare by guaranteeing integrity, registering consent, providing provenance on the decisions impacting care, and auditability of the decision-making process.

There is also a major application for wearables/IoT devices. Collecting and analyzing this data and supplementing it with patient-provided/physician registered information from the EHRs can offer further insight into patient/population health. In this case, blockchains help record provenance of the data.

There are cases where owing to privacy concerns, PII data is stored off-chain and only metadata pertaining to a patient's record is stored on the blockchain.

Thus, blockchain serves as the trust fabric to bind together data from disparate silos, thereby providing a collaborative platform for health information exchange.

SPACE APPLICATIONS

Oftentimes AI is used to analyze the model of operations represented by the digital twin no matter where the real facility is located, even if the equipment is in space.

Use of blockchain could provide a new form of metadata (where the data came from, who processed the data, at what time were the last edits done) to the Earth Observation (EO) data that enables integrity verification on EO imagery products. Blockchain can be used to provide a secure and traceable data exploitation mechanism.

From there, it is possible to:

+ Provide an essential integrity verification mechanism to all EO data products that are archived. Third party auditable truth will likely also increase the demand for future use of the data.
+ Provide auditability and proof to any third party in case of dispute over EO data product delivered through the platform.

In general, blockchain can be used to protect the integrity on any space-based observation data or other related data.

RESEARCH CHALLENGES

Making AI systems and applications more trustworthy is an active field of research, containing several fundamental open questions [4]. Based on published literature, the intersection of this field with blockchain research seems to be narrow, as the technologies are rather orthogonal. Since there is no fundamental difference between AI systems and conventional software systems, all the typical blockchain uses apply to AI systems: proof-of-existence (timestamping), proving the integrity of data and logs, software revisions, provenance, etc. The emergence of significant novel research challenges from the convergence of AI and blockchain, including but not limited to blockchain based federated learning, remains to be explored.

Meanwhile, to avoid overhyped view on the subject, critical thinking should be applied to AI and blockchain related claims published in even peer-reviewed scientific literature, as the review process is apparently often too shallow and overly enthusiastic speculations end up being presented as facts.

CONCLUSION

The primary area of convergence for AI and blockchain stems from the fact that blockchains can be used a data exchange platform and to store anchors to training data, training access and models. Blockchain can also be used for tracking provenance and ownership rights of data.

While AI and blockchain both remain active fields of research, there is a lot of learning ahead of us. It is critical to sustain research to identify how the synergy of these two technologies shapes up in the future.

REFERENCES

1. Independent high-level expert group on AI set up by European Commision - Ethics Guidelines for Trustworthy AI
2. [Breck 2019] Eric Breck, Neoklis Polyzotis, Sudip Roy, Steven Euijong Whang, Martin Zinkevich. Data validation for machine learning. 2019. https://www.sysml.cc/doc/2019/167.pdf
3. [Renngli 2019] Cedric Renggli, Bojan Karlaš, Bolin Ding, Feng Liu, Kevin Schawinski, Wentao Wu, Ce Zhangh. Continuous integration of machine learning models with ease.ml/ci: towards a rigorous yet practical treatment. 2019. https://www.sysml.cc/doc/2019/162.pdf
4. [Arnold 2019] M. Arnold, R. K. E. Bellamy, M. Hind, S. Houde, S. Mehta, A. Mojsilović, R. Nair, K. Natesan Ramamurthy, D. Reimer, A. Olteanu, D. Piorkowski, J. Tsay, and K. R. Varshney. FactSheets: Increasing Trust in AI Servicesthrough Supplier's Declarations of Conformity. 2019. https://arxiv.org/pdf/1808.07261.pdf
5. [Weld 2018] Daniel S. Weld. The Challenge of Crafting Intelligible Intelligence. 2018. https://arxiv.org/abs/1803.04263
6. [Kim 2018] Hyesung Kim, Jihong Park, Mehdi Bennis, Seong-Lyun Kim. On-Device Federated Learning via Blockchain and its Latency Analysis. 2018. https://arxiv.org/abs/1808.03949
7. https://github.com/tensorflow/models/blob/master/research/object_detection/README.md
8. http://cocodataset.org/
9. https://github.com/tensorflow/tfjs-models/tree/master/coco-ssd
10. https://modelzoo.co/
11. https://resources.wolframcloud.com/NeuralNetRepository
12. https://modeldepot.io/