[[1]](#footnote-1)

GDPR Compliant Public Blockchain

using AI Smart Contracts

Andy Ho, An Nguyen and Jodi Pafford

*Abstract*— Blockchain is a highly attractive new technology, due to its security and decentralization features, in a time where concern for consumer data protection is on the rise. The immutability of blockchain through its decentralized framework is an appealing option for conforming to regulations protecting consumer data privacy such as the European Union’s (EU) General Data Protection Regulation (GDPR) and similar legislation proposed in the United States (US), however companies needing to conform to regulations will likely need to keep costs as low as possible and solutions simple. Many of the proposed solutions include storing personal data on a private/permissible blockchain while storing the identification on a public blockchain – aka hybrid model. We will propose and describe a novel solution allowing for all data to be stored on a public blockchain utilizing smart contracts and artificial intelligence (AI). Smart contracts can be designed to automate actions based on conditions set while AI provides predictions and classification values. This mechanism will give individuals full-control of their personal data while providing a method of compensation between consumers and businesses. This will be GDPR complaint while fully using all the security and decentralize features of a public blockchain.

# INTRODUCTION

As the fourth industrial revolution ushers our society into an age of robotics, artificial intelligence, quantum computing, etc… - individuals will become completely dependent on technology in their daily lives. The fuel to power this revolution is the personal data of individuals – identification, transactions, lifestyle choices, medical history, work history, social media presences, and more. There are many companies where people are the product (i.e. Facebook, Instagram, Twitter) and anticipate more to arise in the coming years. The protection and control of a person’s personal data has become the top priority for companies while consumers continue to provide their sensitive information willingly.

In the last few years, there has been massive data breaches resulting in personal data being illegally obtained and used for malicious purposes. Target, a US retail company, had their system compromised via a Supply Chain attack where customer’s credit card numbers were stolen. Facebook, an American online social medial and social networking service company, was in the spotlight for the Cambridge Analytica scandal for misusing users’ data. Equifax, one of the nation’s largest credit reporting companies, massive data breach potentially leading to millions of identify theft. There are several other examples which makes it increasingly more important to find new security measures to protect (and give full-control to individuals) personal data.

Blockchain technology has attracted interests from a wide span of industries; mainly due to its ability to operate in a decentralized fashion [1] and its high degree of anonymity. First introduced by Satoshi Nakamoto in 2008 in a whitepaper for a cryptocurrency called Bitcoin, a blockchain utilizes a digital ledger of transactions where all participants edit in a secure way and is shared over a distributed network of computers [2]. The blockchain has an append-only structure, which helps it protect old data against modification or deletion [3]. In order to make changes, all the nodes present in the network must evaluate, verify, and match the transaction information; if the majority of the nodes agree a new block is added to the chain [2]. Considering several recent highly publicized data breaches raising public concern, and in the face of looming legislative changes in the United States [4, 5], after the implementation of the General Data Protection Regulation (GDPR) [6], this extra-secure framework is an attractive one.

Some possible challenges for companies trying to conform to GDPR-like legislation include the “Right to be forgotten” where citizens are given strict control over their personal data [7]. Conforming to a GDPR-like regulation and securing the data are two dynamically opposed paradigms that must be reconciled with each other. While this currently affects users in the EU, there are ramifications for US companies who have presence and transact in the EU.

One solution that has been presented includes implementing a blockchain-traditional database hybrid where user data is stored on a traditional database and modifications are recorded on the blockchain [8, 7]. Alternatively, another solution is a blockchain with the ability to forget has been proposed. This proposal uses the pruning features of traditional blockchains like Bitcoin or newer smart-contracts like Ethereum to remove blocks in a traceable way [3]. Although these solutions satisfy GDPR, it trades away the full security and decentralization features of a public blockchain.

Our contribution will be to introduce a novel conceptual design to insure GDPR compliant while using a completely public blockchain leveraging artificial intelligence in tantum with smart contracts.

This paper is organized in the following: I. Introduction, II. Blockchain which will provide background and key features, III. General Data Protection Regulation which will define the articles that are most conflicting to blockchain, IV. Current Related Work which is peer research on a couple current work around to have blockchain be GDPR compliant, V. Conceptual Design: Artificial Intelligence Smart Contracts on Public Blockchain which is our novel approach on how to leverage smart contracts and artificial intelligence for a GDPR compliant public blockchain, VI. Conclusion and Future Work.

# Blockchain

Blockchain technology has the potential to redesign how computational resources interact in an automated and decentralized society. This technology was invented by a person (or group of persons) named Satoshi Nakamoto in 2008 to be used as a public transaction ledger for a cryptocurrency called Bitcoin [9]. On a high-level, the ledger is a self-governing list of records called blocks [9] which are linked cryptographically using a hash algorithm. Each block is then connected to the previous block containing a timestamp and the transaction data. Here’s a high-level diagram to illustrate the transaction flow:

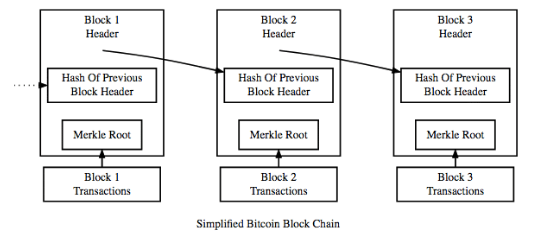


Figure 1. Blocks linked to one another

**2.1 How does Blockchain Works?**

The following will use Bitcoin, a cryptocurrency, to explain how blockchain technology works – it is important to note that this framework can be used in non-finance related applications as well. Owner A wants to send his Bitcoin to Owner B – which in reality assigns Owner B’s identification to that specific transacted Bitcoin. For this to take place, the transaction is inserted as a ‘block’ which is then broadcasted to the peer-to-peer network for verification. If successful, the transaction will be recorded in a public ledger. These ‘blocks’ are all linked to one another (hence, Blockchain) in a linear and chronological sequence with every block containing the hash of the previous block [10] – shown in Figure 1. Refer the following diagram so visualize the flow of how a transaction is processed for Bitcoin:

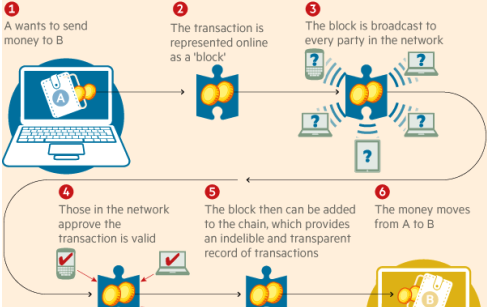


Figure 2. Flow chart of Bitcoin Transaction

**2.2 What is a Cryptographic Hash Function?**

A cryptographic hash function protects sensitive information, either at rest or in transit. In the case of Blockchain, it prevents the double-spend problem (definition: the act of using the same coin more than once) from occurring with the use of public-key cryptography [11]. A transaction is first initiated by future owner of the cryptocurrency by sending his public key to the current owner. The cryptocurrency is then transferred by the digital signature of a hash – the public keys (i.e. the assigned address of the cryptocurrency) are stored in the blockchain. In the case of Bitcoin, it utilizes a SHA-256 hash function which take an input of a random size and produces an output of a fixed size – pre-image resistant. What makes SHA-256 powerful is that it is nearly computationally infeasible to reconstruct a given input from the output value.

**2.3 Immutability of Blockchain**

Immutability is one of the most important (and defining) features of blockchain. As explored in the section above, blockchain is a one-way hash function making the records irreversible without community consensus. This eliminates reconciliations and establishes trust in the system [12]. Only the owner of the record who has the proper credentials could make changes to the records.

**2.4 Industry Applications of Blockchain**

Since 2008, blockchain technology has been considered for many use cases outside of cryptocurrency (i.e. finance related). The following are industries where blockchain technology could be applied:

1. Insurance: With the use of a blockchain application called ‘Smart-Contracts (reference), insurance claims could be processed without the aid of an adjuster or physical inspection. Users would provide the information and the smart-contract would determine if it satisfies the criterions prior to distributing funds to the insured.
2. Internet of Things (IoT): There are new security vulnerabilities since these devices are sending and receiving data. Blockchain would add an additional layer of security insuring only the owner is receiving access and information of his IoT devices.
3. Healthcare: The encryption blockchain provides would be of importance to medical records, prescriptions, and supply management – offering extreme privacy.
4. Voting: Voter fraud could be prevented since a vote recorded in the blockchain would be immutable and would provide an audit trail if there is evidence of tampering. Since each voter would have his own lock and key, authentication of user would virtually be impossible to fake.

Although created for the financial world, the implications of Blockchain technology can affect applications in wide range of areas outside of finances. One of the main purposes of blockchain is to eliminate the need for an intermediary to verify and process a transaction – making it decentralized and distributed. The main objective is that the blockchain establishes a new standard and model by creating a distributed consensus in a digital world [10]. In order to accomplish this, the designers made blockchain immutable and/or resistant to data modification – any change would require all subsequent blocks to be altered. Since these public ledgers are managed by peer-to-peer network [10], mass collaboration governs whether a transaction can be changed. This headwind hits directly at one of the central tenants of GDPR [13].

As one would expect, the supporters of blockchain believe the advantages of it outweigh the regulatory issues – the inverse applies for those who hold data privacy in higher regards. Before moving on to evaluating possible solutions that can satisfy both worlds, lets dive into GDPR.

# General Data Protection regulation

**2.1 What is GDPR?**

*“The improvement in substance is that there’s far more transparency under the new rules, which means that you will have more detailed information policies about what your data are processed for, which purposes if they are given to others, and there will be also in general more possibilities to get a view of which data are there about you. And you have new rights like data portability and the right to be forgotten. So, it will be far easier for consumers to control their personal data.”*

- Jan Philipp Albrecht, member of the European Parliament and ‘father’ of the GDPR [13]

GDPR was created from the European Commission to reform data protection across the European Union in order to make Europe ‘fit for the digital age’ [14]. The origins of what is now known as the GDPR began in 2012. All organizations in the member-states across Europe, including those who have dealings with businesses in Europe must adhere to the GDPR EU framework. The GDPR was approved and adopted in April 2016 but was not enforced until May 25, 2018 [15].

*The key changes in the reform include:*

* *A****single set of rules****on data protection, valid across the EU. Unnecessary****administrative requirements,****such as notification requirements for companies, will be removed. This will save businesses around €2.3 billion a year.*
* *Instead of the current obligation of all companies to notify all data protection activities to data protection supervisors – a requirement that has led to unnecessary paperwork and costs businesses €130 million per year, the Regulation provides for increased****responsibility and accountability****for those processing personal data.*
* *For example, companies and organizations must notify the national supervisory authority of serious****data breaches****as soon as possible (if feasible within 24 hours).*
* *Organizations will only have to deal with a****single national data protection authority****in the EU country where they have their main establishment. Likewise, people can refer to the****data protection authority****in their country, even when their data is processed by a company based outside the EU. Wherever****consent****is required for data to be processed, it is clarified that it has to be given explicitly, rather than assumed.*
* *People will have easier****access to their own data****and be able to****transfer personal data****from one service provider to another more easily (right to data portability). This will improve competition among services.*
* *A****‘right to be forgotten’****will help people better manage data protection risks online: people will be able to delete their data if there are no legitimate grounds for retaining it.*
* *EU rules must apply if personal data is****handled abroad****by companies that are active in the EU market and offer their services to EU citizens.*
* ***Independent national data protection authorities****will be strengthened so they can better enforce the EU rules at home. They will be empowered to fine companies that violate EU data protection rules. This can lead to penalties of up to €1 million or up to 2% of the global annual turnover of a company.*
* *A new****Directive****will apply general data protection principles and rules**for****police and judicial cooperation****in criminal matters. The rules will apply to both domestic and cross-border transfers of data* [14]*.*

The GDPR at its core is powerful and needed, however, the application of it interferes with the premise behind Blockchain Technology. In 2012 when the European Commission first introduced the GDPR, blockchain was not a known word and the GDPR idea was initially focused on cloud services and social networks.

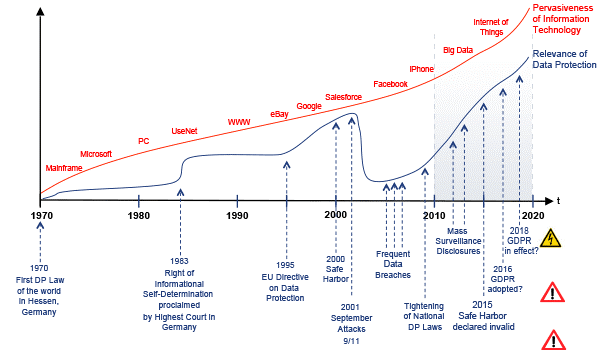


Figure 3. A brief history of the General Data Protection Regulation by Wilhelm (2016) [16]

**2.2 Implications of GDPR for Blockchain**

The table below, summarized from The Journal of The British Blockchain Association, summarizes the implications related to blockchain and GDPR. [13]

|  |  |  |
| --- | --- | --- |
| **GDPR Article/Recital** | **Implications** | **Topic** |
| Art. 4(1), 6(4), 32/Rec. 26 | Can PD be stored on a blockchain or must be off-chain? The connection between pseudonymized and anonymized data and the data subject. | **Personal data on the blockchain** |
| Art. 6 | Six reasons can be used to comply with lawful processing, and a data sharing agreement can be recorded on a BC | **Lawful Processing in the EU/Consent** |
| Art. 17, 17(1), (a,b), 6(1)(b,f)/Rec. 69 | Can data on a blockchain be deleted in accordance  to the RTBF and what would happen if not – could  the functioning principle take over that allows for  specific interpretations of the GDPR, as BC is at its  core designed not to be compliant to the RTBF. | **Right to be forgotten (RTBF) and functioning principle** |
| Art. 25/Rec. 78 | BC runs counter to data minimization, storage limitations and a clearly determined data controller, raising the question whether it is in line with ‘Privacy by Design’ (PbD). Privacy risks of entire IT-architecture, including BC. Solutions could be Enigma or differential privacy or future more secure BCs. | **Privacy by Design versus blockchain core features** |
| Art. 26(1)/Rec. 79 | Private versus public BC and the accountability of a (joint) data controller. | **Accountability of data controller** |

Table 1. Implications related to Blockchain and GDPR.

Article 4

Article 4 of the GDPR defines personal data. The definition is very broad which complicates its interpretation with the use of blockchain [17].

Article 6

Article 6 discusses the consent that must be given. Such consent must have already undergone a thorough academic and practical discourse [18]. This means that consent must be “freely given, specific, informed and unambiguous”.

Article 17

Article 17 of the GDPR grants EU citizens the ‘right to be forgotten and to data erasure’ at any time upon request. Due to the immutable nature of blockchain, this presents a challenge. This is probably the largest challenge of GDPR and Blockchain.

Article 25

Article 25 of the GDPR discusses handling personal data by the concept of Privacy by Design. Privacy by design is privacy such that it “should be promoted as a default setting of every new IT system and should be built into systems from the design stage” [19]. The blockchain implication is that the data must not be stored in plaintext. GDPR does not provide many details to this and has left it up to some interpretation [18].

Article 26

Article 26 discusses the description of who is responsible. This must be completed in a transparent manner in order to be in compliance, which can be a challenge when there are joint data controllers.

# Current Related Works

**4.1 Introduction**

With the growing interest in blockchains as an information storage system and the growing concern for user data privacy research into designing a blockchain based system that are compliant with privacy laws such as the GDPR have begun. Here we identify two strategies proposed by researchers to answer the “right to be forgotten” clause of the GDPR. One proposal by Farshid et al. is to use a pruning algorithm on a smart contract based blockchain, such as Ethereum, and will be describe in the following section [3].

**4.2 Smart Contract Approach**

Smart contract based blockchains have the additional attribute of being able to execute code. All the nodes and blocks together makeup one instance of a virtual machine. This machine can store all account balances and active codes. Once deployed smart contracts only has write-access and cannot be updated or changed. Smart contract based blockchain is attractive because the virtual machine does not require a transaction history to operate but only the current state of the machine. The researchers made use of the pruning algorithm in two Ethereum implementations to delete as much state data as possible without breaking the functionality of the blockchain. Furthermore, the researchers deleted all historical blocks and logs leaving only the current state active. With these changes it was shown that a five host machines were able to form a network and perform basic transactions. They were able to show that an account can be changed and then the history of the exchange deleted. Contracts were able to be created and then the creation transaction be deleted. Figure 4 is a before and after screenshot of an account balance and creation of a smart contract with their transaction history deleted.

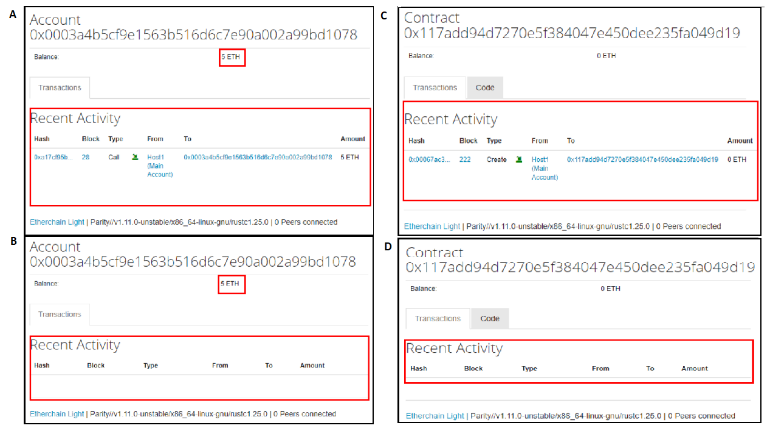


Figure 4. Demonstration of pruning algorithm

There are several limitations to this approach pointed out by the authors. One, no new nodes can be added to the network. This is because the information needed to derive the current state of the virtual machine no longer exist. Two, there is no way to prevent individuals from creating backups of old data before it gets deleted.

**4.3 Hybrid Blockchain Approach**

A different approach is described by Coelho et al [7]. The authors proposed a hybrid system where all “meaningful data” is stored off the blockchain and on a third-party database system. A digest of each instance of data and all transactions performed on the data is stored on the blockchain ledger. This way data can be deleted when requested while at the same time trust in the integrity of the data is proven with the immutability of the blockchain. Figure 5 is a diagram of such a design.

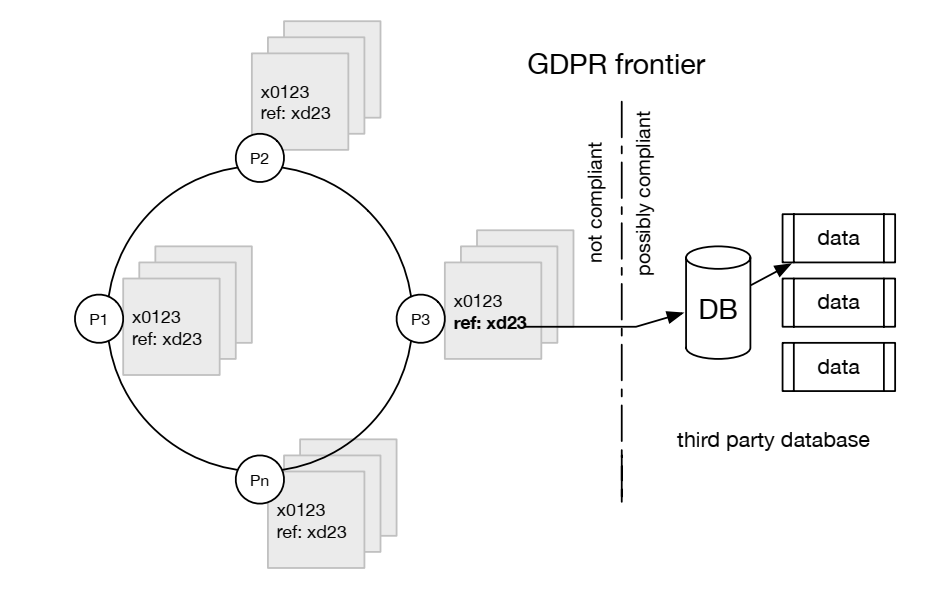


Figure 5. On-ledger/off-ledger hybrid proposed by Coelho et al.

The hybrid system proposed can be integrated into a much larger architecture that uses smart contracts to request permission from the data owners [20]. Figure 6 is a diagram of the ecosystem proposed by Faber et al. using the on-ledger/off-ledger hybrid data storage system. Briefly, the ecosystem utilizes smart contracts to store conditions for data exchanges between users and service providers as well as users and data purchaser. The required permissions allow each member of the ecosystem to interact with the blockchain, contains pointers to the actual data. A second blockchain is used to store hashes of data allowing for the data purchasers to verify the integrity of the data they are accessing. Finally, the data itself is stored off-chain on third party databases.

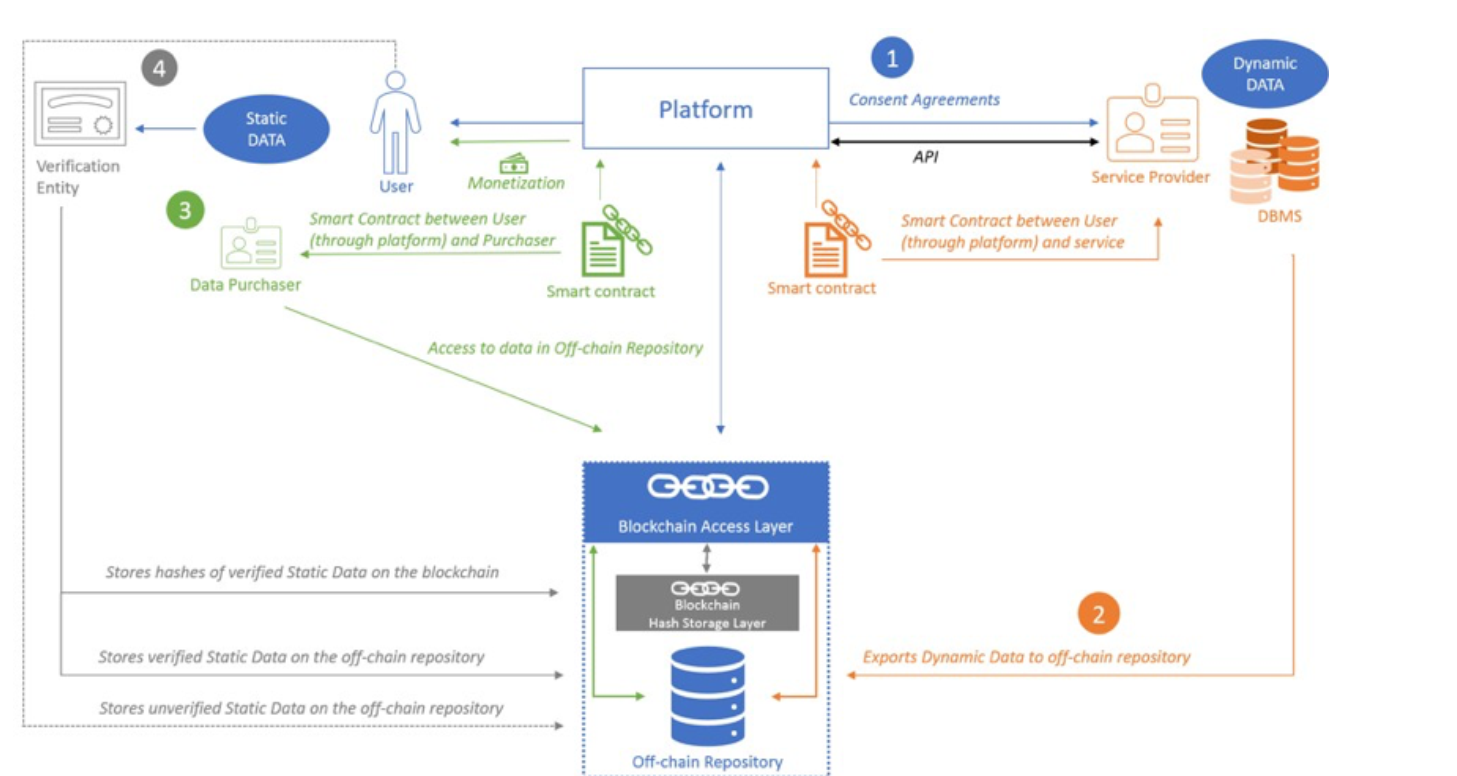


Figure 6. Personal data ecosystem proposed by Faber et al.

**4.4 Conclusion**

Using blockchains to store data is attractive in that a blockchain is immutable and that data is guaranteed to be “un-hackable.” When data storage is moved off the chain it becomes GDPR compliant in that data curators can delete or modify it in accordance with the wishes of the data owner. However, this also opens up avenues for attackers to modify this data. More importantly, it does not take full advantage of the security features of a blockchain compared to that of the conceptual design presented.

# Conceptual Design: AI Smart Contract on Public Blockchain

**5.1 Introduction**

The conceptual design’s primary goal is to make a public blockchain GDPR compliant using artificial intelligence and smart contracts. This novel approach will further enhance the security, protection, and transparency of personal data management for users and requesting entities. As a public blockchain, the database will be fully decentralized with no organizations or entities controlling it. The proposed conceptual design is a foundational building block and is universal for future enhancements from hardware, software, algorithms, and overall data science improvements.

Furthermore, this design will create an ecosystem of data flow, data ownership, and compensation to parties who provide and entities whom make requests. Inversely, as user’s update their personal information on the blockchain, entities will be alerted of the changes and automated events could trigger based on conditions set on the smart contracts. And most importantly to GDPR, users control their data and can request deletion to entities who have stored copies of their personal data. This system would have a full audit trail with time stamps, events, transactions, and a ledger of funds exchanged.

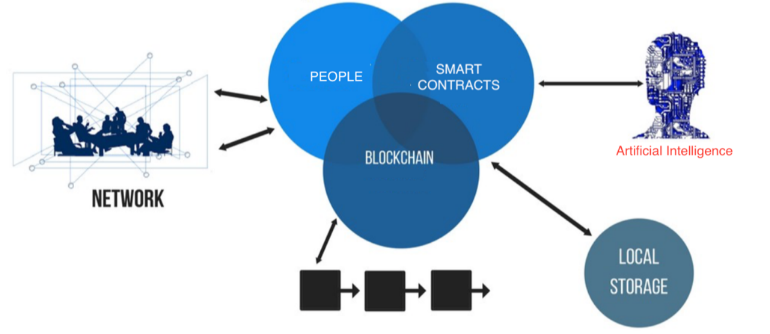


Figure 7. Conceptual Design

**5.2 System Overview**

As illustrated in figure 7, our contribution involves several components and technology. The following are the key stakeholders in our concept:

1. People: Users who are providing their personal data while performing creation, updates, and deletions.
2. Local Storage: Entities who are requesting copies of each user’s personal data and storing on their own private database.
3. Blockchain: A public blockchain that is fully decentralized, transparent, pseudonymity, and governed by collaborators.
4. Smart Contracts: A pre-defined program of conditions (set by users) on who, how, where, why, and when their personal information can be accessed. Also includes the compensation associated to every action (i.e. creation, update, deletions).
5. Artificial Intelligence: Predictions and classifications will provide insights for smart contracts to trigger and execute events.
6. Network: End-to-End communication on a fully decentralized network. Users will self-govern with no interference from 3rd parties – completely open source.

**5.3 Smart Contracts**

A smart contract facilitates, verifies, and executes the pre-defined conditions for each user’s data. Since a smart contract has no ability to think or reason, as its main purpose is to follow a set of instructions, our conceptual design provides artificial intelligence to in essence make it an intelligent contract. Figure 9 is a workflow diagram illustrating the stages of a smart contract.

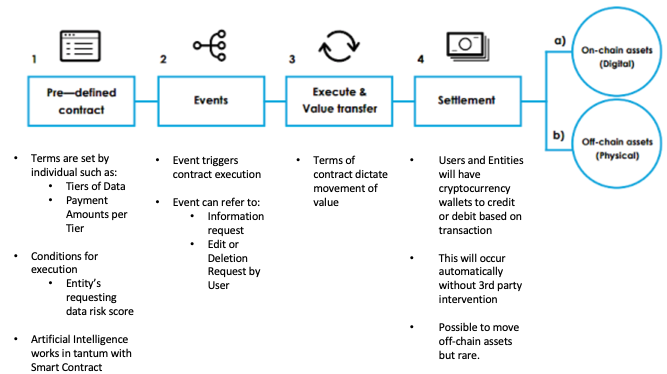


Figure 8. Smart Contract Workflow

**5.4 Tiers for User Data**

In the pre-defined contract section, users would provide the following three (3) tiers of information via a mobile and/or web application service a form and assign their monetary or cryptocurrency values to each:

1. Tier 1: Full Name, Email Address, City, and State
2. Tier 2: Address, Date of Birth, Gender, Income, Phone Number
3. Tier 3: SSN, Driver’s License Number, Passport Number, Credit Score

The tiers can scale and more information such as employment history, medical history, credit card history, IoT devices can be added as well.

**5.5 Sample Code for API to Blockchain**

This information would be transmitted to the blockchain via a JSON file and would add to the block with each update – along with timestamp of transaction. Figure 9 shows a portion of the code for the conceptual public blockchain.

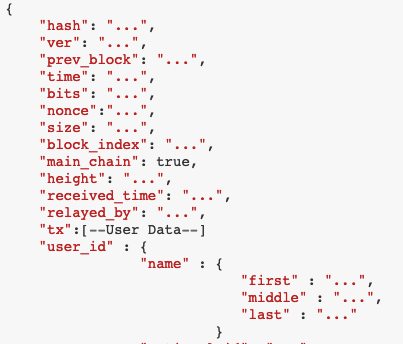


Figure 9. Blockchain Sample

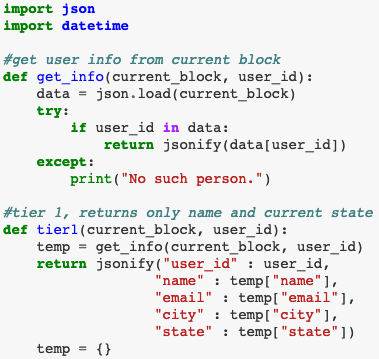


Figure 10. JSON Tier 1 Code

The API would insert or update users’ personal data as shown in figure 10. Since this is a public blockchain, deletion would not be allowed but users to have the right to update their information to incomplete or better yet, have a status indicating user data is not available to be requested.

**5.6 Artificial Intelligence**

The novel concept hinges on the data mining, machine learning, and data insights produced by the artificial intelligence (AI). The possibilities are endless, and applications of the models/algorithms produce will grow, leading to a new industry and opportunity. The following are a few ideas on how AI can serve to bring more intelligence to the smart contracts:

1. Predictive Risk Score of Entities (i.e. Trustworthiness, Security, Avg. Response Time to Requests)
2. Classification of Entities Industry (i.e. Retail, Financial, Social)
3. Categorization of Users and Their Data
4. Recommendation Engine on Compensation bi-directionally
5. Verification of Information for All Parties

**5.7 Analysis of AI**

To demonstrate a possible interaction between smart contracts and artificial intelligence, we imputed an existing data set on San Francisco restaurant ratings to fit the parameters for this paper. This data set was suitable to illustrate how both supervised and unsupervised machine learning can provide insights to guide the execution of the smart contracts pre-defined conditions. The end result will be exploratory data/data mining analysis, feature selection, and possible algorithms that are relevant to scoring whether an entity requesting data is trustworthy enough to access multiple tiers of personal information. There are many machine learning algorithms that can be applied but we have recommended a few that seem to be a good fit for the purposes of the paper.

**5.8 Exploratory Data Analysis**

The data set has over 50,000 observations and 14 features to score San Francisco restaurants quality and trustworthiness. For this paper, we edited the data set to perform the feature selection and relevant algorithms that could lead to a cybersecurity trustworthiness index score (CTSI). It is important to note this is a fictitious data set to demonstrate the calculation and interaction between AI and the smart contract. After cleaning the data set, we proceeded to perform exploratory data analysis (EDA) to observe any possible patterns or missing data or outliers. This will serve to guide the feature selection and help select optimal machine learning algorithms.

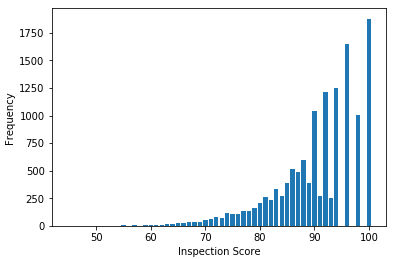


Figure 11. Histogram of Inspection Scores

From this EDA, data scientist can pass on to AI the observations and begin categorizing businesses to learn which features could be significant in its impact (both positive and negative) to the CTSI. Using the figure 10, a mean inspection score can be the maker AI registers to score the variance for any particular business.

Referring to figure 12, the color meaning: Green – Low Risk, Blue – Moderate Risks, Red – High Risks, and Grey - Unknown. There is a cluster of businesses in the northeast end of San Francisco but overall, the businesses labeled ‘High Risk’ are spread throughout the map. Further exploration is required and possibly more data to clarify the businesses marked as ‘Unknown.’

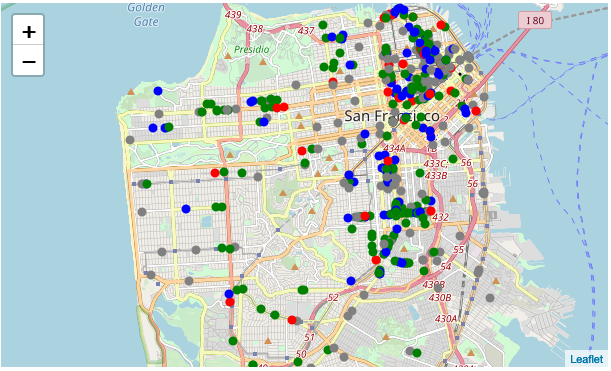


Figure 12. Risk Map of Businesses

In figure 13, it shows that over half of the businesses from the data set have ‘low risk’ while the remaining are moderate to high risk.

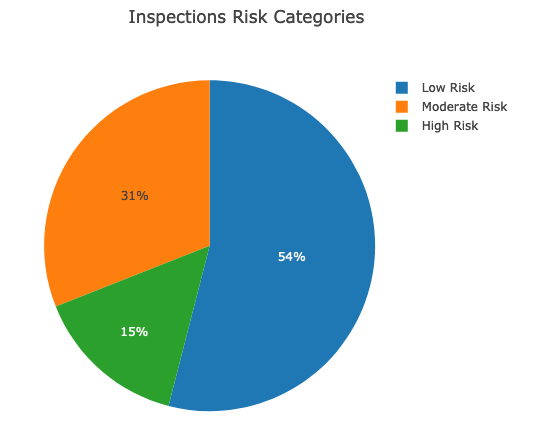


Figure 13. Stratification of Risk

The exploratory data analysis is one that is critical in deciding which features and machine learning algorithms can be used on the data set. There is a huge emphasis on critical thinking to understand who, what, where, and why of data source as it holds large ramifications on inference and generalization.

**5.9 Feature Selection**

Using an unsupervised machine learning such as Random Forest (RF) would provide AI the ranked features that are significant for the CTSI. After performing on the data set, the following features were heavily weighted for the CTSI:

1. Inspection\_Score

2. Violation\_Description

3. Business\_Postal\_Code

These three features would be included in future calculations of CSTI for every entity requesting user data. If the requesting entity does not have any of the three attributes present, then a notification would be sent rejecting their request. If present, then a CSTI would be calculated and passed on to the smart contract as a parameter resulting in approval for the various tiers of personal data.

**5.10 Relevant Algorithms**

Machine learning (ML) is a subset of Artificial Intelligence and is the scientific study of algorithms and statistical models. The goal is to effectively perform specific tasks by relaying on patterns and inferences. To date, ML has four types of categories of algorithms:

1. Classification: Assigns something to a discrete set of possibilities using Decision trees, K-Nearest Neighbor, Logistic Regression, Random Forests, Support Vector Machines
2. Clustering: Assigns set of observations into subsets that are similar in some aspect using Centroid, Density, Distribution, Hierarchical, K-Means, Mean Shift.
3. Regression: Predicts a numerical value using Linear Regression, Lasso, Ordinary Least Squares, Polynomial, Ridge, Splines, Stepwise.
4. Reinforcement Learning: Determines the ideal behavior within a specific context to maximize performance.

For the San Francisco restaurant data set, classification and regression algorithms were utilized to arrive at the CSTI model. In the ‘Feature Selection’ section, random forest algorithm was applied to find the most important features from the data set to include in the model.

**5.11 Feasibility**

The current technology and sophistication of AI is still not a stage ready for automation. Network enhancements such as 5G will allow for better performance and reduce latency. Edge computing for IoT devices could reduce the load on computational resources while gathering more data points for AI to learn.

One of the major roadblocks to a public blockchain is the time required to verify and receive consensus on user data prior to an event executing. Another is adoption and confidence in a fully decentralize system. Both users and entities must be conditioned to have faith that a public blockchain is secure, transparent, and reliable.

Furthermore, AI still requires many human intelligences to instruct and provide it guidance at its current state. This could lead to vulnerabilities within the code and errors in AI’s predictions and classifications.

# Conclusion and future work

In this paper, we proposed a conceptual design for a public blockchain to be GDPR compliant using artificial intelligence and smart contracts. The ideal solution would have a fully decentralized and democratized system that affords everyone full control of their personal information, holding businesses and entities accountable, and compensating all parties involved accordingly – being GDPR compliant.

From our research, there is still a great deal of development required in hardware, software, and legislation. For future work, we would like to explore Natural Language Processing to analyze string variables from data sets and Deep Learning algorithms such as Neural Network to add to AI’s pattern recognition and predictions.

Lastly, there is still the issue of data integrity from users. Because deletion is not allowed on a public blockchain, users can update their information to incorrect data which would lead to unintentional fraud since entities would be compensating users for inaccurate data. We would like to use artificial intelligence to verify information is accurate and legitimate.

# Bibliography

|  |  |
| --- | --- |
| [1] | K. Christidis and M. Devetsikiotis, "Blockchains and Smart Contracts for the Internet of Things," *IEEE Access,* vol. 4, pp. 2292 - 2303, 2016. |
| [2] | S. Singh and N. Singh, "Blockchain: Future of Financial and Cyber Security," in *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, Noida, India, 2017. |
| [3] | S. Farshid, A. Reitz and P. Roßbach, "Design of a Forgetting Blockchain: A Possible Way to Accomplish GDPR Compatibility," *Hawaii International Conference on System Sciences |,* pp. 7087-7095, 2019. |
| [4] | National Conference of State Legislatures, "Data Disposal Laws," National Conference of State Legislatures, 04 January 2019. [Online]. Available: http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx. [Accessed 1 January 2019]. |
| [5] | National Conference of State Legislatures, "Security Breach Notification Laws," National Conference of State Legislatures, 29 September 2018. [Online]. Available: http://www.ncsl.org/research/telecommunications-and-information-technology/data-disposal-laws.aspx. [Accessed 31 January 2019]. |
| [6] | THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION, *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (GDPR),* L 119 ed., 2016. |
| [7] | F. Coelho and G. Younes, "The GDPR-Blockchain Paradox: A Work Around," in *Workshop on GDPR Compliant Systems*, Rennes, France, 2018. |
| [8] | C. Molina-Jiménez, I. Sfyrakis, E. Solaiman, I. C. L. Ng, W. Meng, Wong, A. Chun and J. Crowcroft, "Implementation of Smart Contracts Using Hybrid Architectures with On-and Off-Blockchain Components," *ResearchGate,* pp. 1-12, 2018. |
| [9] | S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," *www.cryptovest.co.uk,* 2008. |
| [10] | M. Crosby, Nachiappan, P. Pattanayak, S. Verma and V. Kalyanaraman, "BlockChain Technology: Beyond Bitconin," *Applied Innovation Review,* no. 2, pp. 6-19, 2016. |
| [11] | T. Aste, P. Tasca and T. D. Matteo, "Blockchain Technologies: foreseeable impact on industry and society," *IEEE Computer,* vol. 50, no. 9, pp. 18-28, 2017. |
| [12] | M. Pilkington, "Blockchain technology: principles and applications," in *Research Handbook on Digital Transformations*, Northampton, MA, Edward Elgar Publishing, 2016, pp. 225-253. |
| [13] | S. Schwerin, "Blockchain and Privacy Protection in the Case of the European General Data Protection Regulation (GDPR):A Delphi Study," *The Journal of The British Blockchain Association,* vol. 1, no. 1, pp. 1-75, 2018. |
| [14] | European Commission - Press release, *Commission proposes a comprehensive reform of data protection rules to increase users' control of their data and to cut costs for businesses,* Brussels: European Commission, 2012. |
| [15] | Trunomi, *EU GDPR COMPLIANCE WITH TRUNOMI – ARTICLE SUMMARIES & SOLUTIONS,* GDPR, Whitepapaer, 2017. |
| [16] | E.-O. Wilhelm, "A brief history of the General Data Protection Regulation," International Association of Privacy Professionals, 2019. [Online]. Available: https://iapp.org/resources/article/a-brief-history-of-the-general-data-protection-regulation/. [Accessed 13 3 2019]. |
| [17] | N. Kramer, "Blockchain, Personal Data and the GDPR Right to be Forgotten," 17 April 2018. [Online]. Available: https://www.blockchainandthelaw.com/2018/04/blockchain-personal-data-and-the-gdpr-right-to-be-forgotten/. [Accessed 13 March 2018]. |
| [18] | C. Wirth and M. Kolain, "Privacy by BlockChain Design: A Blockchain-enabledGDPR-compliant Approach for Handling Personal Data," *roceedings of the 1st ERCIMBlockchain Workshop 2018, Reports of the European Society for SociallyEmbedded Technologies,* pp. 2510-2591, 2018. |
| [19] | B.-J. Koops and R. Leenes, "Privacy regulation cannot behardcoded. A critical comment on the‘privacy by design’ provision in data-protection law," *International Review of Law, Computers & Technology ,* pp. 37-41, 2013. |
| [20] | B. Faber, G. Michelet, N. Weidmann, R. R. Mukkamala and R. Vatrapu, "BPDIMS: A Blockchain-based Personal Data and Identity Management System," *Proceedings of the 52nd Hawaii Interantional Conference on System Sciences,* pp. 6855-6864, 2019. |

1. MSDS 7349 – Network and Data Security – Spring 2019 [↑](#footnote-ref-1)