**Conceptual Design: AI Smart Contract on Public Blockchain**

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.

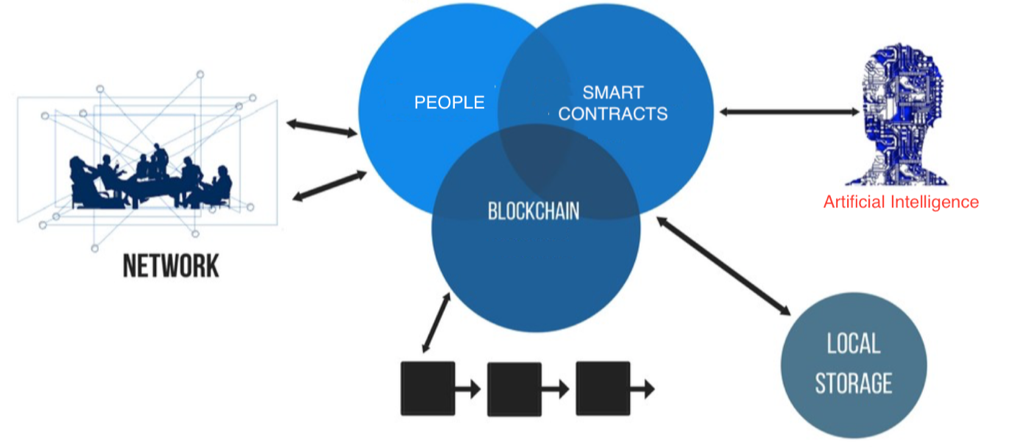


Figure 5: Conceptual Design

**System Overview**

As illustrated in figure 5, 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.

**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. The following is a workflow diagram illustrating the stages of a smart contract:

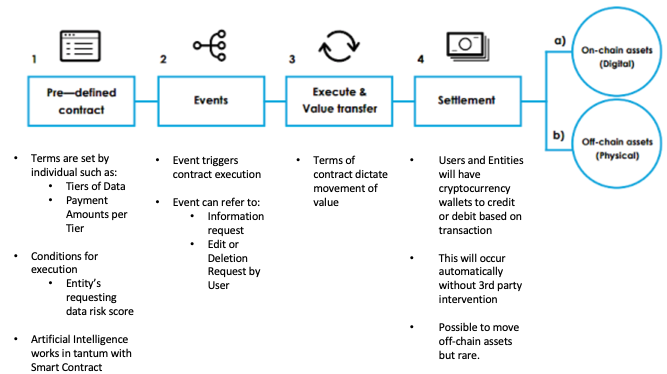


Figure 6: Smart Contract Workflow

**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:

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

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.

**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:

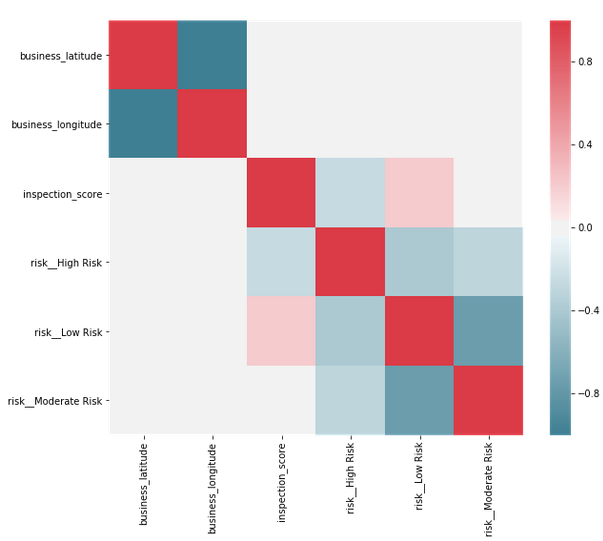
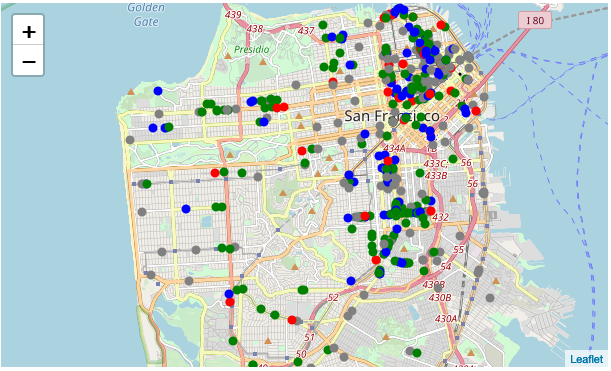
1. Predictive Risk Score of Entities (i.e. Trustworthiness, Security, Avg. Response Time to Requests)
2. Classification of Entities Industry (i.e. Retail, Financial, Social)
3. Categorization of Users and Their Data
4. Recommendation Engine on Compensation bi-directionally
5. Verification of Information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

**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.

**Exploratory Analysis**

The data set had over 50,000 observations and 14 features to score San Francisco restaurants quality and trustworthiness. For our purpose, we edited the data set to perform the feature selection and relevant algorithms that could lead to a cybersecurity trustworthiness index score (CTSI). Here are a few graphs from our EDA:

From a simple EDA, AI can begin categorizing businesses and learn which features could be significant in its impact (both positive and negative) to the CTSI. In this case, AI would learn that businesses in more populated area have had more violation incidents then those in less populated. More data science would have to be performed but this demonstrates how EDA would be recorded during a supervised machine learning process.

**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.

**Relevant Algorithms**

<any suggestions on ML algorithms that can be used>

**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. This could lead to vulnerabilities within the code and errors in AI’s predictions and classifications. And from our analytics, another area called Natural Language Processing would be required to catalog and process values to arrive at relevant insights since we anticipate many of the data sets provided to AI will be string variables.