Article

Not peer-reviewed version

Missing Links: Current Trends and Future Potential in the Application of Blockchain Oracles

Cedric Heidt *, Philipp Sandner, Marc Anders

Posted Date: 6 May 2024

doi: 10.20944/preprints202405.0218.v1

Keywords: smart contracts; automation; fintech; oracles; process optimization; web3

 doi:10.20944/preprints202405.0218.v1

Article

Missing Links: Current Trends and Future Potential in the Application of Blockchain Oracles

Cedric Heidt *, Philipp Sandner and Marc Anders
Frankfurt School of Finance & Management
* Correspondence: c.heidt@fs.de

Abstract: Oracles have been suggested as a vital piece of infrastructure necessary to allow blockchain-based smart contracts to reach their full potential by interacting with real-world data. This paper begins by providing an approachable entry point to the oracle problem and the current landscape of existing solutions. From there, we provide an overview of published research to narrow down unexplored research directions and derive questions for interviews. Finally, the results of 15 unstructured expert interviews are presented and discussed. The focus here is the industry landscape of oracle projects, opportunities for innovation, and challenges needing to be overcome.

Keywords: smart contracts; automation; fintech; oracles; process optimization; web3

1. Introduction

As blockchain technology matures, the need for reliable and efficient methods to integrate real-world data into decentralized applications becomes increasingly vital. Blockchain oracles serve as intermediaries facilitating this crucial information flow from off-chain sources to on-chain smart contracts. This paper aims to contribute to current research on blockchain oracles, hybrid smart contracts, and the interplay between off-chain and on-chain elements. The paper is positioned to provide easy access to readers who may be unfamiliar with the topic while offering insights with respect to potential future extractable value.

1.1. Blockchain: A Trust Machine

For years now, blockchain technology has been making waves. At times, being proclaimed a key driver for the next industrial revolution [1,2]. At others, condemned as nothing more than a useless gimmick powering a gigantic Ponzi scheme [3]. Regardless of the public and market sentiment toward the technology, there are certainly reasons it has received so much attention beyond speculation.

In its most simplified form, a public blockchain is a shared append-only ledger that enforces certain rules while allowing anyone read and write access. This distributed digital ledger allows users to deterministically achieve consensus on a common reality without the need for centralized parties to keep records and govern the process [4]. Blockchain technology was first introduced in Bitcoin, though the term "Blockchain" was never explicitly mentioned in the original whitepaper by Nakamoto [5]. Bitcoin was designed to function as a peer-to-peer electronic cash system with a native, independent currency and monetary system tied to participation in the maintenance of the network and processing of transactions [5]. For the first time in history, there was a globally standardized currency that anyone could use, but nobody could monopolize control over due to the fundamental design of blockchain consensus protocols.

While this in itself is a historic achievement, researchers interested in the technology quickly realized other potential use cases in which such a system might yield significant benefits and potentially revolutionize how we transact, exchange value, and keep records. Early examples mentioned by Tapscott [6] include the standardized recording of land titles, providing financial services to all regardless of factors such as location and access to legal documents, decentralized sharing economy platforms, self-sovereign digital identity, automated royalty payments for digital content, and even technology-enabled systems to support and improve governmental functions. In general, blockchain technology primarily enables two functions: (1) digital scarcity enabling value transfer and (2) decentralized processing and validation of transactions immutably stored in a chronologic ledger; the

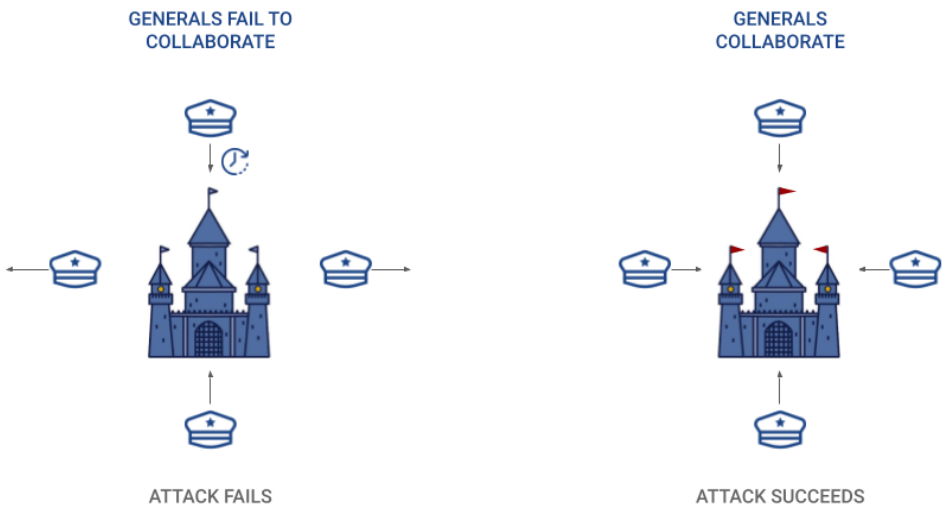
 doi:10.20944/preprints202405.0218.v1

emphasis concerning more advanced applications beyond basic party-to-party asset transfer is on the latter. While digital scarcity certainly plays a role, decentralized computation and execution of transactions can be used to create systems and protocols for various use cases. The innovation here is that neither trust among participants nor the involvement of a central governing party is necessary for anyone to participate in a system or use an application.

Decentralized computation via a smart contract network allows for guaranteed neutral execution of pre-defined logic. This is useful as it removes the need for trust when engaging in even complex transactions. Decentralized computation is practical when building so-called Decentralized Applications (dApps). User-facing applications utilizing a blockchain in its back-end. Currently, the primary use for dApps is in finance, primarily on the Ethereum blockchain [7]. Decentralized finance (DeFi) is an umbrella term for decentralized exchanges, lending protocols, derivatives, insurance, and more. All logic for such applications is written ahead of time in the form of a smart contract and published on a blockchain. Once published, the code cannot be changed but it can be audited by all parties. The logic outlined in the contract is executed when a party interacts with the contract [7]. This allows parties to engage in complex transactions without the need to trust one another - the operation is trustless [8]. There is no need for a bank, a government, a notary, an escrow agent, or any other middleman whose role traditionally is to help overcome issues of trust. This has the potential to drastically streamline and disintermediate a wide variety of transactions, even beyond finance.

Before going further into detail about the blockchain's role as a trust machine, it is essential to understand decentralization and how it is achieved and maintained. At the foundation of every blockchain is the consensus protocol - in other words, how network participants agree on a common truth to be immutably recorded on the ledger [9]. Depending on whether the blockchain is public, allowing anyone to participate in consensus, private, allowing only allowlisted nodes to participate, or a hybrid system, different methods are used to ensure network integrity. Underlying these different approaches is the need for Byzantine fault tolerance.

The term Byzantine fault comes from the so-called Byzantine Generals Problem, an allegory describing the fundamental challenge of ensuring a certain minimum percentage of actors behave as expected. The analogy used to describe this problem involves multiple generals planning to attack a fortress. The battle is won if all generals coordinate the attack and follow through. If, however, certain generals fail to participate in the coordinated attack or retreat instead of attacking, the battle is lost. Generals cannot be expected with absolute certainty to communicate honestly. Messengers between them may be intercepted, the other party may inject false messengers, and generals may outright lie [10].



Note. Illustration of the Byzantine generals problem. Own work.

Figure 1. Byzantine Generals Problem.

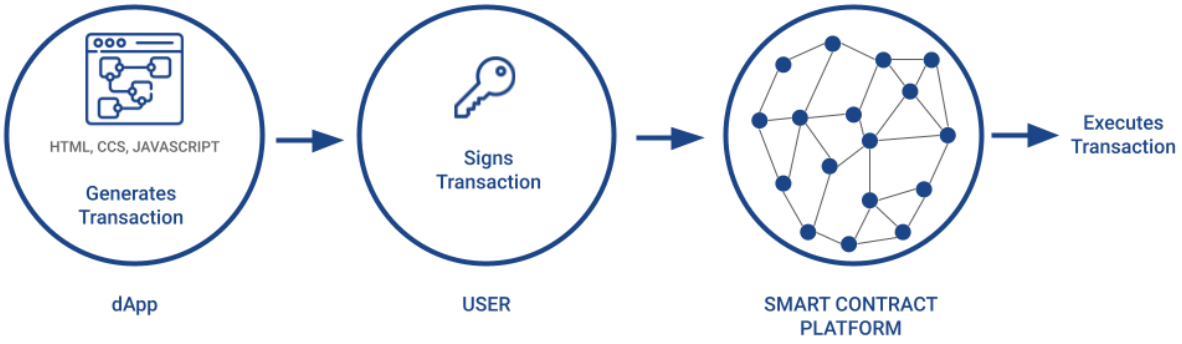
 doi:10.20944/preprints202405.0218.v1

The two leading permissionless consensus protocols, meaning anyone can participate and participants cannot inherently be trusted, used in public blockchains to overcome this issue are proof of work (PoW) and proof of stake (PoS). In proof of work, network participants use computing power to solve hash-puzzles [11] while in competition with each other. Economic incentives ensure that participants use their computing power to secure the network rather than attack it [5]. Proof of stake works very similarly, though it does not use computational power but rather direct financial incentives. Participants in PoS consensus commit funds as collateral to contribute to transaction validation. These foundational protocols ensure the network's neutrality or decentralization [12]. There are other approaches to consensus, some of which build on a certain element of trust achieved by only allowing known actors to participate, also known as permissioned blockchains. This, of course, also creates a limited network that may be easier to target [13].

Byzantine fault tolerance of blockchains allows for a provably fair and independent network able to store and process data without the need for a central authority. In short, blockchains are useful when trust between two parties wishing to transact does not exist, which is quite common in a digital economy.

1.2. The Oracle Problem

Bitcoin was originally designed as a decentralized peer-to-peer cash system able to not only execute transactions between parties but also to utilize a native currency with automated monetary policy [5]. While extreme price volatility [14] and high transaction fees [15] have hindered Bitcoin's adoption as a digital cash system, as demonstrated when Steam removed Bitcoin as payment method in 2017 [16], some argue Bitcoin has found use as a form of "digital gold" - a relatively uncorrelated digital commodity [17]. With Bitcoin's continued proof that a decentralized blockchain system can reliably secure value, new projects began emerging. Ethereum, for example, took the concept further by adding Turing completeness - the ability to execute computer code [18]. At Devcon1 in 2015, Ethereum co-founder Vitalik Buterin used the allegory of Ethereum being a smartphone on which Bitcoin is merely one single application [19]. This type of Turing complete Blockchain is also sometimes called a smart contract platform. Smart contract platforms have the ability to execute series of "if-then" statements, allowing for logic to be written onto the network. Using a traditional web front-end and connection to a smart contract platform, developers have the ability to build applications with decentralized execution [20]. Such applications have the potential to offer a tremendous amount of value as they replace the need for trust with cryptographic guarantees.



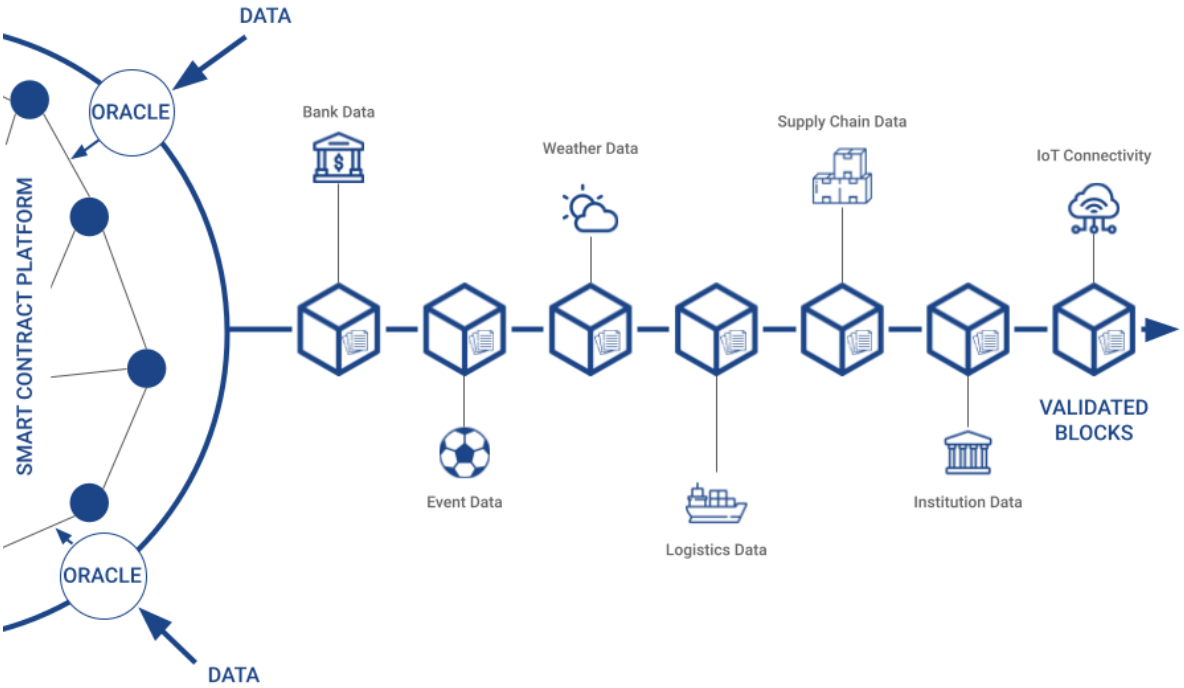
Note. Figure showing dApp transaction execution. Own work.

Figure 2. dApp Architecture.

Blockchains have been operational for around 15 years now, and while the promise of a world with a reduced need for intermediaries has attracted some attention, mainstream adoption, and real-world use cases are still few and far between, at least when compared to the full potential of the technology. While this may largely be attributable to scaling problems found in many blockchain networks, difficult user experience journeys, and an inherent lack of understanding of the technology [21], an argument

 doi:10.20944/preprints202405.0218.v1

could also be made that decentralized applications that could make a real positive difference in every day life have not yet been built. Whereas simple decentralized applications are already available and quite successful by some metrics, for example, decentralized peer-to-peer lending platforms and decentralized exchanges [22], more advanced applications that would offer value in the daily life of most average people are not. When developing smart contracts, there is a limit as to what can be done natively within a certain blockchain network. Such networks are isolated systems by design. This provides excellent levels of security and reliability. However, it also makes the development of smart contracts that rely on external data challenging to build. The potential of decentralized execution is limited without the ability to observe and interact with off-chain systems. Hence, the question of how developers may provide data access to their smart contracts without potentially introducing major security vulnerabilities arises. Is there a way to minimize centralization and maximize data integrity in data provision? This challenge is sometimes called the oracle problem [23]. Consequently, third-party services that enable data availability are referred to as oracles.



Note. Visualization of the oracle problem and use cases. Own work.

Figure 3. The Oracle Problem.

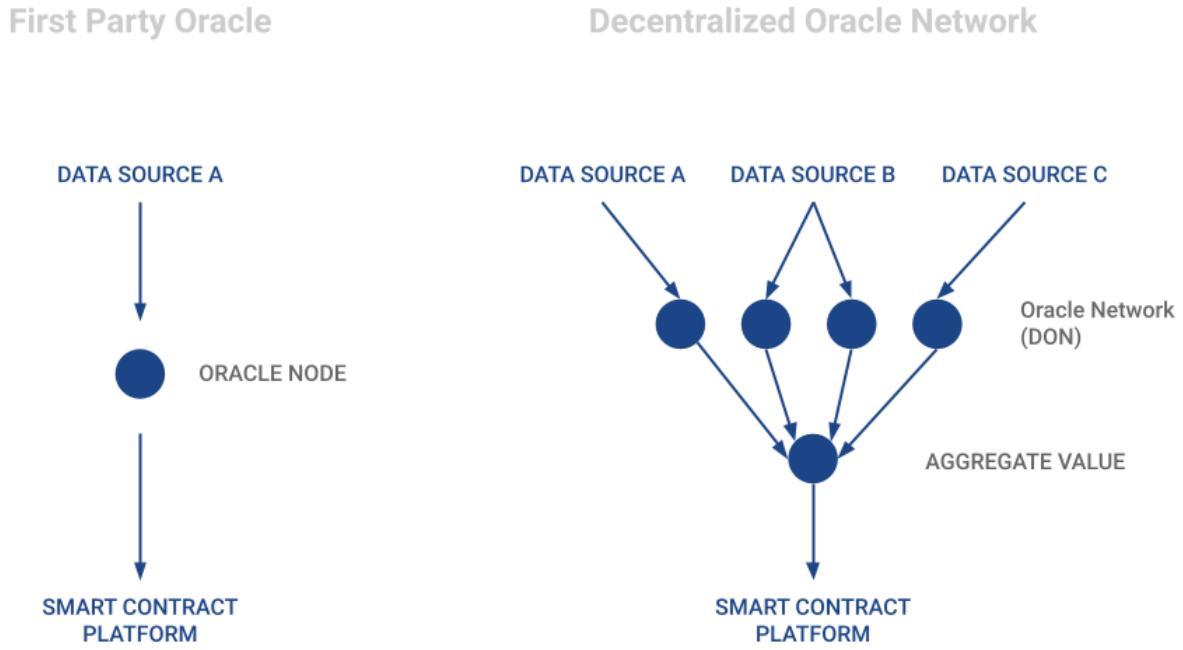
Currently, there are several approaches to allow smart contracts to access blockchain-external data. The most primitive way to enable data access are centralized oracles. Centralized oracles provide data from a single external source and are operated by a single entity. Such oracles are very easy to implement and often purpose-built. However, being centralized, they also present a single point of failure. If the data is corrupted or manipulated, it may lead to significant ramifications [24]. Alternatively, several projects are working on implementing decentralized oracles. Decentralized oracle implementations are the focus of this paper. Among them are Chainlink, Band Protocol, API3, and many others. A brief summary of how these projects tackle the Oracle problem follows.

Chainlink is considered the industry leader, going by various metrics, including the total number of integrations [25]. Chainlink is focused on third-party oracles. Theoretically, anyone can set up and operate a node that makes data available on-chain. Node operators are not automatically considered trusted. Economic considerations are used to create trust, with dishonest node operators being punished. A very similar approach is used by Band Protocol (Band, n.d.-b), though using their own blockchain network. API3, on the other hand, relies on trusted first-party oracles to solve the oracle problem (API3, n.d.). Whereas Chainlink and Band Protocol rely on a network of nodes to fetch

 doi:10.20944/preprints202405.0218.v1

data from external data sources, e.g., APIs, API3 enables API providers to operate their own oracle node directly. There are, of course, many other approaches to solving the oracle problem. While this paper primarily focuses on decentralized third-party oracle networks, particularly Chainlink's implementation, all approaches will be taken into consideration.

Third-party-based oracle systems approach to overcoming the oracle problem is called a decentralized oracle network, DON for short. As briefly explained above, a decentralized oracle network comprises a committee of nodes [26]. DONs ingest data through a connection between the oracle node and an Application Programming Interface (API) provided by one of many data providers. Different nodes within a single DON gather data from multiple independent sources, and a median of all the different data reported by the various nodes is made available to the smart contract.



Note. Centralized oracles versus decentralized oracle networks. Own work.

Figure 4. Decentralized Oracle Network Data Aggregation.

Aggregation and the ability for anyone to operate a node within a DON or to create a new DON, provide a high level of reliability, redundancy, and a degree of decentralization. Nodes also sign the data sent to the network, proving which node was the one to send it. Even though the data is aggregated off-chain, these signatures remain attached to the data and are also sent on-chain. Since this signature is stored on a blockchain, there is a transparent and permanent historical record of the performance of a certain node. Reputation systems allow smart contract developers to filter nodes by certain metrics [26]. This introduces competition for node operators and is an additional incentive to keep nodes honest. In addition to data sourcing and delivery, oracle networks may offer various services, such as trustlessly triggering actions in a smart contract, creating and providing a source of randomness, and more. Oracle functions are discussed in part V of Theory.

1.3. Research Direction

After the introduction, the theory section delves into the importance of data in both Web2 and Web3. It explores the potential applications of hybrid smart contracts and their technical foundations. Additionally, it presents a comprehensive overview of the predominant oracle protocols in use today, along with their functions and key stakeholders. The theory chapter aims to provide readers unfamiliar with smart contracts and oracles with the necessary foundational information.

The overarching goal of this paper is to contribute to the available research surrounding blockchain oracles by providing insight regarding the following three guiding research questions. (1) How do

 doi:10.20944/preprints202405.0218.v1

stakeholders position themselves within the competitive landscape, and what are the criticisms and advantages of their approaches? (2) How mature is the industry in its current usage, and what is its future outlook? (3) What opportunities are enabled by the industry, and what tangential businesses arise from it? These questions will be further refined and positioned to fill gaps in current research identified in the literature review section of this paper. Following the literature review and the methodology chapter, the insights gained through 15 expert interviews are presented.

2. Theory

In 2006, British mathematician Clive Humby said, "Data is the new oil." [27]. In the years since then, huge companies with data as a core part of their business model have climbed up the market capitalization rankings of major stock markets worldwide. The internet has enabled new possibilities for data sourcing, transfer, and use cases involving data. With new technologies such as AI and blockchain, the significance of data is posed to only increase. Decentralized computation and digital currency promise to enable a wide variety of new use cases in the same way the internet has, potentially reshaping the entire data industry. This section aims to explore the data economy of today, provide an in-depth summary of how oracles work, why this topic is significant, as well as the current state of the technology. Furthermore, the landscape of oracle protocols and stakeholders in oracle systems are explained.

2.1. Data Economy

The importance of data in business has increased considerably in recent years, fundamentally changing how businesses function and generate value. For organizations to succeed in the modern digital era and acquire a competitive edge, data has long been considered a crucial strategic asset [28]. Data is being used by businesses to improve customer experiences, streamline operations, spur innovation, and make optimal decisions [29]. The digital economy runs on data, which enables businesses to open up new income sources and provide value for their consumers and shareholders. Using services and exchanging goods produces information that is stored, transferred, and used. This subchapter will examine the present situation in today's data economy, as well as highlight the current value of data as a strategic asset.

Considering that some of the largest companies in the world by market capitalization are technology companies with a strong emphasis on data in their business models, data is clearly of high strategic importance. Apple, Microsoft, Alphabet, Amazon, and Meta, in particular, would likely never have reached their current dominant positions without the strategic integration of big data in their business models. Be it generating and monetizing data by selling it to advertisers, or using data to increase revenue in some other way, for example, by up- and cross-selling, or using data as a competitive advantage by creating walled garden ecosystems surrounding it. This phenomenon is nothing new and has been clearly identified for many years now. According to a study by Davenport and Harris [30], businesses that effectively leverage data are more likely to achieve superior performance than those that do not. The study found that data-driven companies outperformed their competitors in terms of profitability, productivity, and market valuation. In addition to technology companies, businesses across all industries are recognizing the importance of data in the context of their operations. They leverage data to optimize processes, improve customer experiences, and make informed decisions. For instance, retail companies use data to personalize their marketing efforts and increase customer loyalty [31], while healthcare providers use data to improve patient outcomes and reduce costs [32]. Even Netflix's popular television series "House of Cards" was originally greenlit based on data analysis that revealed viewers of the original BBC version of the drama, also like films starring Kevin Spacey or produced by David Fincher [33]. Netflix knew the show would be a hit without seeing a single scene.

The ability to collect and analyze vast amounts of data has opened up new opportunities for businesses to create value and gain a competitive edge in the market. While this is hardly surprising, and the term "big data" is commonly used, there are many different ways for companies to participate

 doi:10.20944/preprints202405.0218.v1

in and profit from the data economy. The following briefly summarizes some of the largest stakeholders and their activities.

2.1.1. Infrastructure

The most basic layer of the data economy is infrastructure. This layer's stakeholders are primarily large technology companies, including Amazon, Google [34], Microsoft, IBM, and Oracle, among others. These companies offer a range of products and services that help organizations manage and analyze their data effectively. AWS, a part of Amazon that offers cloud-based computing and storage services, brought in $80 billion in revenue in 2022, a large percentage of which came from data-related services, including data warehousing, analytics, and machine learning [35]. The Cloud business of Google, which provides comparable services, produced $26.2 billion in revenue in 2022 [34]. The Azure cloud platform, which offers various data-related services like data analytics, machine learning, and artificial intelligence, contributed significantly to Microsoft's $198 billion in sales in 2022 [36], making it another prominent player in the big data market. The cloud division of Oracle, which provides data-related services, including database management, analytics, and machine learning, generated over $10 billion in revenue in 2022 [37]. These companies provide the backbone for the data economy, primarily storing and processing vast quantities of data.

2.1.2. Data Brokers & Aggregators

Companies that collect and sell data are often called data brokers or aggregators. These companies specialize in collecting, analyzing, and selling large amounts of data. Many of these companies focus on different types of data [38]. In general, they can be separated into companies focused on personal data, and companies focused on non-personal data. Some of the largest data aggregators focused primarily on personal data include Acxiom, Experian, Equifax, TransUnion, Oracle Data Cloud, and Neustar, among others [39]. These companies collect data from various sources, including public records, social media, online activity, and consumer transactions, among others. They then package this data into profiles or segments, which are sold to other companies for targeted advertising, lead generation, and other marketing purposes. On the non-personal data side, companies such as Bloomberg and Thomson Reuters specialize in collecting and selling financial data, including market data, news, and analytics, to financial institutions, investors, and other businesses. Other companies, such as IHS Markit, specialize in collecting and selling data related to industries such as energy, automotive, and technology. In addition, some companies specialize in collecting and analyzing sensor data from devices such as Internet of Things (IoT) sensors, smart home devices, and industrial equipment. These companies include firms such as Intel, IBM, and Siemens, among others. They collect and analyze sensor data to help businesses optimize their operations, improve efficiency, and identify new business opportunities.

2.1.3. Data Users

Finally, there are the companies and organizations that purchase the generated data. The key customers in the data economy who purchase data from brokers can vary depending on the type of data being sold and the specific industry involved. Different organizations, with different strategic objectives, purchase different types of data. Commonly quoted examples include marketers and advertisers, using consumer data to identify potential customers and create targeted advertising campaigns [40], healthcare companies, which purchase data including patient demographics, medical history, and clinical results in an effort to improve patient outcomes and reduce costs [32], government agencies, and researchers, to name a few.

More interestingly in the scope of this paper is the use of data by financial institutions. These organizations use data to assess risk, prevent fraud, and optimize investment portfolios, among other things. One concrete example of this is BlackRock's Aladdin. Short for Asset, Liability, and Debt and Derivative Investment Network [41], Aladdin is a technology platform that ingests data and provides

 doi:10.20944/preprints202405.0218.v1

investment management, risk management, and operational tools to institutional investors [42]. As of 2020, Aladdin managed $21.6 trillion in assets [43]. In order to power Aladdin's capabilities, BlackRock purchases a variety of data from different sources. Some examples of data that is purchased for Aladdin include market data, for example, market prices, trading volumes, and other indicators, economic data, including indicators such as GPD, inflation, and employment, fundamental data on individual companies, such as financial statements, earning reports, and analyst estimates. Other purchased data types include satellite imagery, social media sentiment analysis, and credit card transaction data [44]. This data is used to gain insights into specific industries or companies and to identify potential investment opportunities. Once acquired, the data is processed and analyzed using machine learning and other quantitative methods.

AI and machine learning is an industry with great promise that relies on large quantities of data to train and improve their models. The more data that is available, the better the AI [45]. Machine learning has made tremendous progress in recent years in the areas of autonomous navigation [46], generation of media [47], efficiency improvements in industry, and more, promising to disrupt a plethora of different sectors, creating a tremendous amount of value in the process. However, some issues related to the data economy are highlighted particularly well in this area of generative AI. An example of such an issue is the question of copyright in generative AI. When an AI model is trained with images from certain artists and learns to replicate their style and patterns, should the artist who created the source material be rewarded, and to what extent? Other issues in the data economy relate to privacy, security, the monopolization of data, and more. Ultimately, advancements in the area of artificial intelligence have the potential to transform the way we work, live, and conduct business, largely through automation. Not just in manufacturing and industry but also in finance, digital technology, and much more. Spanning industries and trillions of dollars of economic activity. As such, it is important to approach this technology thoughtfully and responsibly.

While data has become a vital resource in almost any type of industry, we may be only just scratching the surface. Data is critical in automation, especially in combination with other technological advancements such as additive manufacturing, blockchain, IoT, and others. Important in the scope of this paper is the overlap between data and decentralized computation and value settlement.

2.2. Data in Web3

In 1997, computer scientist, legal scholar, and cryptographer Nick Szabo published a paper titled "The God Protocols" (1999). In this paper, he describes his vision for decentralized, cryptographically guaranteed fair systems that are more efficient and secure than traditional systems - Mathematically Trustworthy Protocols. The use cases for such protocols proposed by Szabo include decentralized currency systems, dispute resolution, contract enforcement, identity, and voting [48]. Such systems, in theory, have the potential to reshape large parts of the economy by increasing efficiency and reducing the need for centralized authorities and intermediaries. In 1998, Szabo proposed a decentralized currency system named Bit gold [49]. The basic idea behind Bit gold was to create a decentralized system for creating and managing a scarce digital asset, which could be used as a medium of exchange and store of value without relying on a central authority. However, Bit gold was never implemented (Van Wirrum, 2018). Over the years, many projects have attempted to fill the various proposed functions of a god protocol. Some specialized in one functionality, others in universality, composability, and interoperability. Most notably Bitcoin, which began operating in 2009 [50]. Bitcoin was originally designed as a peer-to-peer cash system, including a decentralized currency and settlement network [5]. In July of 2023, Bitcoin is currently the decentralized network with the most nodes [51] and has a market capitalization of over $500 billion [52]. Bitcoin proved that such a decentralized protocol is possible in practice. With one of the proposed use cases of such a god protocol achieving at least some success, developers became inspired to find ways to build even more capable systems.

Today, there are thousands of protocols. Nonetheless, Szabo's vision of the potential of a god protocol has not yet been fulfilled. That is not to say that no progress has been made. Smart con-

 doi:10.20944/preprints202405.0218.v1

tract platforms, such as Ethereum, have the fundamental functionalities necessary to implement decentralized contract enforcement, dispute resolution, identity systems, voting systems, and more [53].

Ethereum, for example, is not simply a decentralized ledger that stores account balances but is instead more fittingly described as a distributed state machine [54] as explained in the Ethereum developer documentation:

> Ethereum's state is a large data structure that holds not only all accounts and balances, but a machine state, which can change from block to block according to a predefined set of rules, and which can execute arbitrary machine code.

This combination of memory and executive capabilities is often also summarized as the Ethereum Virtual Machine or EVM for short. The EVM functions as a decentralized computer composed of a worldwide network of participants able to execute arbitrary functions [54]. In practice, it is already possible to build decentralized financial service applications, organizational structures governed by voting, identity and reputation protocols, and much more. However, one limiting factor to Ethereum's functionality is data availability. While developers have a toolset of functions, Ethereum is an isolated system that does not natively allow for outside connections. While this separation is essential for security and stability, it limits the type of applications for which the EVM may be used.

This is a problem since much of digital innovation is combinatorial in nature. On one hand, the open-source ideals found in the blockchain space are a tremendous advantage. On the other hand, new services often rely on data feeds in the form of Application Program interfaces and Micro Services that integrate specific functionalities not available within the closed system of the Ethereum Virtual Machine. In Web2, this is not an issue. Organizations may easily leverage external APIs. Ridesharing applications, such as Uber for example, do not use infrastructure built from the ground up, instead using pre-built APIs. Often a GPS API for location data (e.g., MapBox), an SMS API for messaging (e.g., Twilio), and a currency API (e.g., BrainTree) for payment processing [55]. These are external services not maintained by or directly accessible by organizations; an interface is necessary. It is no different regarding applications built to utilize a decentralized network for computation. The problem regarding connectivity and reliability that arises in decentralized systems is often referred to as the oracle problem.

2.3. Significance

In order for applications built to run on a decentralized computation network to reach their true potential, they need to be aware of and able to communicate with external systems [56]. Being able to ingest real-world data, such as market prices, allows smart contracts to use them as triggers for an "if-then" statement on-chain. Having the ability to send messages from on-chain to blockchain external systems would allow them to trigger events such as sending payments off-chain or sending instructions to physical devices, to name just a few examples. In order to enable this type of functionality a piece of middleware is needed. This piece of infrastructure is referred to as an oracle [56]. Smart contracts that use oracles for their processes are also called hybrid smart contracts since they combine off-chain and on-chain data and computation [57].

While use cases of blockchain oracles are very diverse and found across many different industries, a commonly cited example is that of parametric insurance, as it highlights not only the functional significance of oracles but also the benefits of cryptographic guarantees and decentralization. Functionally, a decentralized parametric insurance application consists of a smart contract containing the insurance logic and a set of oracles providing the necessary data to evaluate and enforce claims [58]. The case of parametric flight-delay insurance is an excellent example: the insured pays a given fee into a smart contract to insure against the cancellation or delay of a specified flight. At a certain point in time, an oracle is used to query details about the flight, and depending on predefined conditions, the claim is automatically settled. The entire process is automated and executed according to the smart

 doi:10.20944/preprints202405.0218.v1

contract. Neither the insured nor the insurer need to trust each other since no one can interfere with the execution [59]. The insured benefits from transparency, immutability, and speed. So long as the oracle reports the correct data, there is no way for either party not to complete the agreement as outlined in the smart contract. Settlement is automated and happens instantly [58]. The other side also benefits from efficiency gains. Rather than relying on third-party insurers, airlines could directly integrate such an application into existing systems and processes, potentially boosting efficiency, reducing potential legal liabilities, and increasing transparency. While it is possible to employ parametric insurance without smart contracts, decentralized implementations benefit from cryptographically guaranteed execution. Once the hybrid smart contract is active, the agreement will be immutably executed as specified therein [60], making trust and legal protection superfluous.

Flight insurance is just one rudimentary example of a decentralized parametric insurance application. The example of crop insurance better highlights the potential of decentralized parametric insurance. Decentralized parametric crop insurance allows farmers around the world access to standardized insurance against specific weather events threatening their crops. If the oracle reports that there was less than a certain amount of rain where the insured are located, they are automatically paid out drought insurance compensation. This simple concept increases the availability of insurance as the insured need only a cheap phone and internet access to enter into contracts. There is no need for extensive documentation or regional insurance providers [58]. This is significant because in LEDCs, formal agricultural insurance is often not available at all [61], and these regions are also the most vulnerable to extreme weather events; exacerbated by global warming [62]. Furthermore, these regions are also more likely to rely on subsistence farming to support local populations [63], further emphasizing the importance of insurance. In other words, these already vulnerable regions are hit the hardest and hurt the most [62]. According to a prediction by the Intergovernmental Panel on Climate Change, the increasing frequency and extent of extreme weather events, such as heatwaves and droughts, will cut agricultural productivity in Africa by up to 50 percent [64]. Already, agricultural productivity has declined by 34 percent since 1961 due to climate change [64]. Access to modern financial services such as insurance is highly important for LEDCs. Many financial products do not exist for speculative reasons but rather to allow people and businesses to mitigate risks and hedge against unfavorable situations [65]. The scalability, security, and transparency provided through hybrid smart contracts has the potential to allow anyone anywhere access to reliable financial services.

Beyond being better for customers, cryptographically guaranteed self-executing contracts lead to significant efficiency gains for companies through automation, allowing for increased profits or reduced prices for customers. The extent of these efficiency gains cannot be understated. A tangentially related example of the impact of this type of technology can be found in the reinsurance industry. According to a report by PwC, the reinsurance industry alone stands to reduce its expenses by 15% to 25% by automating settlement using smart contracts amounting to $5-10 billion saved yearly (2016).

The insurance sector is far from the only industry likely to be disrupted by implementing hybrid smart contracts. Other examples of where potential use cases are to be found include the financial sector as a whole, supply chain management, gaming, marketing, government, mobility, and manufacturing [66–68]. Everywhere conventionally a middleman or additional process has been needed to establish trust there is potential for disruption [69]. In the financial industry alone, there are countless potential use cases for hybrid smart contracts, as many industry subsections exist purely to facilitate transactions between parties and enable hedging against risk.

While critics might say that cryptographic guarantees and decentralized networks replacing centralized authorities is unnecessary as regulation makes these intermediaries reliable, some cases prove otherwise, underscoring how imperfect a system relying on centralized trusted authorities to function can be, no matter the level of regulation. In the lead-up to the subprime mortgage crisis of 2008, one specific middleman played a significant role: ratings agencies (Bush, 2022). Since 1936, US banks have been legally required to act only on “recognized ratings manuals” regarding the buying and selling of bonds. “Recognized ratings manuals” refers to the information published by the rating

 doi:10.20944/preprints202405.0218.v1

agencies, typically one of the big three agencies: Moody's, Standard & Poor's, or Fitch. In the leadup to the financial crisis, rating agencies gave mortgage-backed securities overly high ratings. When the subprime mortgage market began collapsing, many highly rated securities turned out to be far riskier than investors had been led to think (White, 2010). Some researchers have highlighted the fact that since rating agencies were paid by the companies issuing the securities, they had the incentive to offer high ratings [70]. The 2008 financial crisis led to a total loss of over $2 trillion in global economic growth [71], millions of jobs, and billions in income. All in large part attributable to the failure of a trusted middleman to remain unbiased. An interesting question to look into is which examples of intermediaries we rely on today pose a particularly high risk and whether automating their functions via a hybrid smart contract would be feasible.

2.4. Use Cases

While parametric insurance is a great example case that highlights both the capabilities and significance of oracle-enabled hybrid smart contracts, the benefits apply to various industries. This chapter summarizes prominent use cases, especially those highlighted in academic publications. This is not an exhaustive list; instead, it is a collection of use cases considered especially significant in current discourse and research. It also provides examples of experimental use case implementations from smaller projects and hackathons.

2.4.1. Interoperability, Real World Assets, Perpetuals

One of the most anticipated use cases for oracles is direct interoperability when moving assets from blockchain to blockchain or between off-chain bank accounts and a blockchain. The potential benefits of blockchain interoperability have been the focus of research for several years now [72]. Different methods to enable blockchains to communicate with one another have been developed, though some implementations fall short, as demonstrated by the wormhole platform hack in early 2022 in which an attacker stole around $300M worth of crypto assets from the aforementioned cross-chain protocol [73]. Certain tokens may be natively available on multiple smart contract platforms with a dApp that allows bridging by burning a token on one chain and issuing it on another. Such tokens are in the minority, with one example being USDC issued by Circle [74]. The majority of tokens require workarounds, such as AMM systems or wrapped tokens. AMM systems essentially allow the trustless exchange of tokens, in some cases including tokens across different blockchains, through linked liquidity pools. This process has potential issues concerning security and attracting and efficiently allocating liquidity. Wrapped tokens are created by locking up any crypto assets on one chain and issuing a new token representative of the collateral on another chain. An example is WBTC, a bitcoin pegged token on the Ethereum blockchain. While in some cases, wrapped tokens are an effective solution, they are seen as a workaround rather than a complete solution [72].

Chainlink Labs recently launched its Cross-Chain Interoperability Protocol (CCIP), a standardized protocol for cross-chain communications. CCIP combines the burning of tokens on one chain and minting them on another for tokens that natively have that function, such as USD; for other tokens, it uses a lock and burn function - essentially creating a wrapped token via an oracle node. Theoretically, oracle nodes can potentially make wrapped tokens more secure and trustless. In addition to tokens, CCIP allows the transfer of arbitrary data between chains. An oracle node receives and forwards the data to the destination chain [75]. While CCIP is still very new, it will be exciting to see whether it finds adoption.

Another missing link that may be solved via oracles is the connection of traditional financial intermediaries to smart contract platforms. In June 2023, Swift, "the most prominent standards body in the global banking community [76]" announced that it was exploring connectivity between permissioned blockchains, as may be used by traditional financial intermediaries and the public Ethereum blockchain via Chainlink oracle nodes [77]. The ability to transfer tokenized assets from

 doi:10.20944/preprints202405.0218.v1

a permissioned blockchain to public blockchains would open the door for a significant increase in liquidity on public blockchains.

While tokenization, the process of digitizing real assets, is still in its infancy, it promises to "help create universally accessible, fast, liquid, and transparent investment and financial systems" [78], and potentially enable entirely new types of digital securities. "A Security Token' is a digital and tokenized version of traditional security, and its value depends on the value of the asset, i.e., the value of the ownership that the token represents [79]." Security tokens are particularly relevant when it comes to ownership rights. The benefits of tokenized assets include the ability to raise capital and liquidity benefits due to fractionalization and theoretical global ease of access [78].

Complementary to digital assets tokenized via collateralization, perpetual futures, often referred to as perpetuals or perps, which follow the price movement of specific assets without being collateralized by that asset, have been gaining popularity. Perpetuals are a derivative providing investors with long or short exposure to an asset without needing to hold a specific asset. Perpetuals use fees to incentivize traders to close the gap between future and spot prices. At the moment, perpetuals are almost exclusively used to represent crypto futures but may well be used to provide exposure to off-chain assets in the future [80]. In order to do this, perpetuals protocols must only be able to access off-chain data pertaining to the spot prices of the assets being traded.

2.4.2. Supply Chain

One specific use case thoroughly discussed in current research is in the supply chain field. It has been hypothesized for years now, that the field of supply chain is ripe for disruption by blockchain-based applications [81]. While there have been some experimental implementations of on-chain supply chain tracking applications, none have had the type of disruptive success that some may have expected. One example is Walmart's blockchain-tracked supply chains for Mangos and Pork [82]. These were implemented in 2018 and discontinued in late 2022 [83]. One main problem that is present in on-chain supply chain management applications is that while the blockchain is an immutable way to store and make data accessible if the information that is put into the system is incorrect, there is no added value of using such a system. As Caldarelli [84] point out:

> As information on the blockchain is immutable but not necessarily true, without a trusted third party to verify the data to be inserted, the details provided should not be considered any more trustworthy than those contained in a legacy database.

Additional barriers to adoption may be found in cost and caused by supply chain fragmentation and outsourcing. Nonetheless, experts agree that adopting blockchain-based oracle-supported supply chain tracking applications can lead to more efficient, sustainable, and transparent supply chains [56].

2.4.3. Other

Additional cases of applications relying on verifiable data inputs to smart contracts can be found across a variety of sectors. In healthcare, for example, there is the potential to share data, such as patient records, between healthcare service providers, leading to efficiency improvements and likely also improved medical outcomes for patients. Additionally, blockchain-based prescriptions may reduce fraud compared to legacy methods [85]. However, health care is rife with sensitive data, and any mistakes may be disastrous. As in the supply chain field, the critical point is inputting reliable, verifiable data into the blockchain-based system while addressing additional privacy concerns. This trend continues when looking into academia, where privacy and reliability are paramount, for example, when storing academic credentials and transcripts on-chain [72]. Beyond this, examples of value-adding applications have been identified in IoT integrations [86], real estate [87], and beyond. The following chapter will take a closer look at some of the functionalities mentioned that are enabled by oracle protocols. Verifiable oracle data feeds, multiparty computation, and zero-knowledge proofs have been identified as crucial elements in hypothetical blockchain-based applications [86].

 doi:10.20944/preprints202405.0218.v1

2.5. Oracle Implementations & Functions

This section focuses on oracle solutions available today. First, a comparison of different oracle protocol implementations is presented. Next, a closer look at data feeds, inputs, and functionalities that allow for the cryptographically guaranteed automation of various services. An overview of a hybrid smart contract developers toolbox, so to speak.

2.5.1. Oracle Risks

This section aims to explain the current industry standards when implementing oracles. While a centralized oracle is easy to establish for each use case, such implementations introduce vulnerabilities. Smart contracts take the data received from oracles as it is presented. Even if the information is clearly false, it is used as instructed. Hence, while hybrid smart contracts themselves will always predictably act as they are programmed to, the contract will not act as it should if the data source used is corrupted or manipulated; the contract will not act as it should. When using a single centralized oracle there is a high risk of this happening [88]. An example can be found when looking at the decentralized lending platform Compound. In 2020, Compound was sourcing data regarding the market value of assets from a centralized oracle run by Coinbase [89]. Compound allows users to borrow on-chain assets so long as they put up other on-chain assets as collateral. If users are unable to maintain a certain minimum percentage of collateralization, their position is liquidated. On November 26th, the oracle reporting the price of DAI, a US Dollar stablecoin, falsely reported the price of DAI as having spiked to $1.30. As a result, some users were now deemed undercollateralized, and over $89 million ended up being erroneously liquidated [90]. Had multiple oracles connected to different exchanges been used to provide the smart contracts with a median price, this situation could have been prevented. Ultimately, this case highlights the need for a degree of decentralization and redundancy with regard to data provision. Decentralized oracle networks are made up of multiple different independent nodes, all independently incentivized to act honestly. Using decentralized oracle networks instead of single centralized oracles has become the state of the art when building decentralized applications.

2.5.2. Oracle Implementations

Many different projects are working on making data accessible to smart contract developers. The following is a summary of some of the most prominent such projects.

Table 1. Oracle Protocols: Overview.

Protocol	Architecture	Networks	Token	Launch	Team
Chainlink	Multiple[a]	11	Yes	May 2019	400+
Band	Algorithmic	18	Yes	Sept 2021b	32
API3	1st Party	10	Yes	July 2021	33
Pyth	1st Party	20	Yes	Aug 2021	25
RedStone	Algorithmic	39	Planned	Jan 2022	14
Tellor	Optimistic	6	Yes	Aug 2019	12
Witnet	Algorithmic	25	Yes	Nov 2020	10
DIA	Algorithmic	8	Yes	2018[c]	25
Supra	Algorithmic	N/a[d]	Planned	N/a[d]	90
UMA	Optimistic	2	Yes	May 2021	136

[a]Algorithmic & 1st Party; [b]Launch of current version; [c]Unclear whether this is start of development or mainnet launch; [d]Data Feeds not yet live.

Chainlink

Chainlink is the apparent market leader for oracles by crucial metrics such as the number of integrations, value secured, number of DONs, and several others [25]. Chainlink was the first oracle project launching on the Ethereum main net in May 2019 [91]. Currently, Chainlink DONs are

 doi:10.20944/preprints202405.0218.v1

integrated into over 1600 projects, they have delivered over 5.8 billion data points to smart contracts enabling over $6.9 trillion in transactions associated with these smart contracts across 11 different smart contract platforms, including Ethereum, Polygon, Arbitrum, and others [25]. Chainlink DONs are used by many prominent protocols, such as the lending platforms Aave and Compound, the insurance protocols Nexus Mutual and Etherisc, and the Loopring exchange [92]. Chainlink DONs comprise several nodes, each using different data sources. There is a minimum number of nodes that can be set for the data to be used and a maximum deviation from the mean that each node should not surpass. Chainlink employs a combination of a reputation system and staking to ensure nodes act as they should [93]. Chainlink DONs are used to bridge off-chain data and provide off-chain computation.

The team behind Chainlink is the largest in the oracle space, currently consisting of over 400 people [75]. The team is supported by advisors such as Ari Juels (professor at Cornell Tech), Tom Gonser (founder of DocuSign), Eric Schmidt (former CEO of Google), Jeff Weiner (former CEO of Linkedin), Balaji Srinivasan (former CTO of Coinbase) and Dan Boneh (professor at Stanford University and the head of its Applied Cryptography Group) [55].

Band Protocol

Band Protocol was launched on the Ethereum main net in 2020. However, this first version received little attention from developers, and the team abandoned the initial design to build Band v2. Band v2 relies on its own blockchain, where node operators also act as validators for this chain's consensus protocol. Nodes combine data retrieved from external APIs to provide data on-chain. However, the types of APIs that can be connected are limited to freely accessible feeds. Password-protected or paid APIs can currently either not be connected or require the cooperation of API providers [94]. Some off-chain computation features are built natively into the protocol. Band protocol currently has limited adoption and is struggling to take advantage of network effects. The team behind Band Protocol consists of roughly 32 people [95].

API3

API3 was launched in July 2021. Included in the founding team are several former Chainlink employees. API3 hopes to disrupt the oracle space by allowing Web2 API providers to make their data available directly to on-chain applications. Rather than using oracle nodes to collect data from various data sources, API3 focuses on first-party oracle nodes. While this design offers some potential advantages, such as transparency of data sources and vastly lower latency [96], it could be argued that this approach brings some of the concerns of centralized oracles. While the need to maintain reputation will likely deter node operators in this system from acting dishonestly, relying on single centralized parties as data sources may introduce certain risks. Currently, API3 is available on ten smart contract networks. Any API can be made available on-chain via the API3-developed middleware Airnode [96]. The team behind API3 currently consists of roughly 33 people [97]. API3DAO, the decentralized autonomous organization used in some governance functions, currently comprises 6538 members [98].

Pyth

Similar to API3, Pyth has taken a first-party oracle approach focusing on low latency, high confidence market data, primarily price data of currencies, crypto assets, and publically traded equities [99]. Currently, Pyth serves over 250 data feeds across over 20 blockchains. Data is served directly by trading firms, cryptocurrency exchanges, and other financial services providers such as Jane Street Capital, Gemini, Cboe Global Markets, and roughly 80 others [100]. Data providers operate a fork of the Solana blockchain named Pythnet. On this chain, data is aggregated and made available to other networks [100]. Having launched in August of 2021, Pyth is on the newer side of oracle platforms. Prominent users of the Pyth network include the synthetic assets platform Synthetix and Tradingview [100]. Pyth currently has around nine employees [101]. Pyth does not currently have an associated token but is planning to launch one for certain governance functions [99].

 doi:10.20944/preprints202405.0218.v1

RedStone

RedStone primarily provides price data regarding cryptocurrencies, stocks, currencies, and commodities, though this platform offers flexibility regarding data types. For example, there is an oracle reporting the number of Refugees leaving Ukraine, the floor price of certain NFTs, and the prices of grains and livestock. Data is primarily sourced from cryptocurrency exchanges and publicly available APIs [102]. RedStone is focusing on both push and pull-based oracles. In push-based oracles, nodes must "push data onto the chain in certain intervals [103]. In pull-based oracles, data is not continuously delivered on-chain but only when requested. Data is stored on the decentralized storage protocol Arweave and is delivered on a specific blockchain by a network of nodes upon request. These oracles are available on over 39 Blockchains [103]. The protocol was launched in January of 2022 [104], currently has around 17 employees [105], and has 1190 different data feeds [102].

Tellor

Tellor is taking an open approach to oracles. A user can request that any data be reported on-chain for a small payment. Data reporters then compete to provide the requested data to claim the tip. Any holder of the token associated with Tellor may lock tokens in order to be able to dispute data reports. When a dispute is opened, the community has two days to vote on the correctness of the report. If the user who opened the dispute correctly flagged the data, the node that supplied the data is penalized. Otherwise, the user who supplied the data receives the locked tokens of the user who opened the dispute. While the dispute is ongoing, a new value is reported to the requester [106]. Tellor is live on the six most prominent EVM main net blockchains, currently has around 12 employees [107], and initially launched in August 2019 [108].

Witnet

The Witnet oracle operates on its own blockchain secured by a decentralized network of nodes that earn WIT from fulfilling requests for data and participating in consensus. Developers may use Witnet for its cryptocurrency price feeds, as a source of randomness, or to connect to external APIs. There are 248 price feeds to 25 smart contract platforms [109]. Witnet originally launched in November 2020 [110] and currently has around 10 team members (Witnet, n.d.-b). Currently, Witnet fulfills around 10,000 to 15,000 data requests per week, according to figures published by the Witnet Foundation. By relying on its own blockchain, Witner Oracles benefits from increased efficiency since the Witnet blockchain is able to communicate data to all other integrated EVM chains without needing a separate decentralized network of node operators for each one [110].

DIA

DIA oracles are focused on providing market data. They currently have around 70 data sources, of which around 45 originate on various blockchains, while 25 are off-chain data sources. Data feeds include prices of various crypto assets, non-fungible tokens (NFTs), and liquid staked derivatives (LSDs). This data is sourced from centralized exchanges, decentralized exchanges, and NFT marketplaces. In addition to price feeds, DIA offers functionality for verifiable randomness on eight smart contract networks, including Polygon, Moonbeam, Fuse, and others [111]. DIA consists of three layers: data collection, data storage and processing, and data publishing. All layers run on a Kubernetes cluster hosted on the IBM cloud [112]. DIA launched in 2018 [113]. The DIA Association currently has around 25 employees [114], and the DIA DAO currently receives between 10 and 30 votes on recent proposals, indicating a rather small community.

Supra

The oracle protocol implementation created by Supra is called Distributed Oracle Agreement (DORA). DORA is made up of a decentralized network of nodes that is itself made up of smaller, periodically randomized networks of nodes [115]. This provides a high level of security in transferring

 doi:10.20944/preprints202405.0218.v1

data on-chain. As of July 2023, Supra has not yet launched its oracle protocol on mainnet and is only live as a beta testnet, making it difficult to evaluate further factors. In addition to its oracle protocol, Supra offers randomness services separately from data feeds; this feature is live on several blockchain networks. Supra oracles participated in accelerator programs run by the University of California, Berkeley, and Mastercard [116] (Supra, n.d.-a). Currently, their team consists of roughly 90 people [117].

UMA

UMA employs a so-called optimistic oracle. Anyone can request a specific piece of data such as "What was the mean temperature in July?" or "Who won the World Cup in 2021?". Users then post a bond and provide an answer to receive a reward. In most cases, the result is assumed to be valid and accepted. In case the information provided is untrue it can be disputed. In this case, the data provider may lose the collateral they deposited. This approach is relatively novel and specializes in delivering a wide variety of data types. It is unsuitable to serve frequently updating data feeds instead posting individual answers to specific questions. UMA is currently primarily used to resolve outcomes in the prediction market "Polymarket" [118]. UMA currently has 136 team members [119]. As of August 2023, UMA has settled 5700 requests since launching on mainnet in May 2021 [120]. An interesting use case of optimistic oracles is using them as triggers in case some external event is observed, e.g., a protocol getting hacked. One downside of this implementation is responsiveness, as oracles must be given a dispute period before being used in a smart contract.

Kleros

Kleros, "The Justice Protocol", is an oracle focused on determining the truth in subjective cases using crypto-economic incentives and a decentralized network of arbitrators rather than digital data links to objective sources. Kleros focuses on disputes, for example, regarding freelancing, e-commerce, content moderation, etc. [121]. Anyone may become a Juror by staking a certain value. Jurors are presented with cases per their indicated areas of expertise and decide based on voting. Game theoretical models incentivize honest behavior and disincentivize malicious behavior [121]. While this approach differs greatly from the objective raw data-based approach of the oracle platforms mentioned previously, it attempts to fill an important niche in operationally necessary on-chain data. Kleros won the European Commission's Blockchains for Social Good prize and a grant from the French bank BPI France [122]. Today, Kleros has a team of around 32 people [123].

Augur

Augur takes a completely different approach to oracles than the platforms mentioned previously. It is an oracle and a platform for prediction markets. Users may create prediction markets for future events, such as who will win the upcoming presidential election, users may then bet on the outcome of the event using the US Dollar stablecoin DAI. After a certain deadline, the outcome is determined by allowing users to vote. Correct voters are rewarded, while incorrect voters are penalized [124]. In theory, this method not only provides real-world data on-chain but also provides data on the likelihood of future events. Augur was one of the first dApps to launch on the Ethereum smart contract platform in July 2018 (Blockchains could breathe new life into prediction markets, 2018). In November of 2021, control of Augur was handed over to a newly created Decentralized Autonomous Organization (DAO). The DAO, in theory, is meant to democratize and decentralize ownership and decision-making power over the platform by allowing members to vote on actions and fund allocation of funds. The latest completed prediction market listed on their dApp was completed in July of 2021 with no currently running markets. Recently, a new similar dApp called Polymarket has been gaining attention [125].

 doi:10.20944/preprints202405.0218.v1

The Graph

The Graph indexes and serves data from various smart contract platforms via APIs. Rather than bringing off-chain data on-chain, they are focused on doing the opposite in a way that is easy to integrate for Web2 developers [126]. The Graph offers a lot of flexibility regarding data types going far beyond price feeds.

While Chainlink is currently the industry leader and, in many cases, considered the industry standard, there are many relevant approaches to solving the oracle problem. Some specialize in only one function, while others focus on offering a wide toolbox. In this paper, Chainlink's terminology and examples of their implementation will, at times, be used; however, the scope of this paper remains project-agnostic.

2.5.3. Inputs and Functions

The inputs made available to smart contract developers by oracles can be split into two main categories: data and off-chain computation. Data may refer to making off-chain data or data from other smart contract platforms that are not natively interoperable available on a specific platform. Computation refers to secure data generation or processing outside of a blockchain or smart contract network to fulfill needs for functions that are not possible or prohibitively expensive when done natively within a smart contract network.

Data

With regard to the former category, data, there is a wide variety of feeds and categories of data that are made available to smart contract developers. In practice, market data is currently by far the most in-demand data category, as this is necessary for many DeFi applications such as exchanges, lending, and derivatives platforms to operate [127]. The market data category can range from simple spot prices of various on-chain crypto assets to liquidity and volume data of commodities and foreign exchange markets. Various oracle projects support different types of this data. While Chainlink provides great flexibility and supports almost all kinds of such data, other, newer oracle platforms have seen the already maturing market for price feeds and focus exclusively on them, Pyth being an example. Market data availability is a key driver of the decentralized finance (DeFi) ecosystem [127]. DeFi is an umbrella term for decentralized applications offering financial services without relying on intermediaries, such as decentralized exchanges, lending platforms, stablecoin protocols, derivatives, and more [128]. MakerDAO, for example, is a protocol allowing users to lock up various crypto assets to mint the US Dollar stablecoin DAI. The locked value must be higher than the amount of DAI minted and if the value of locked assets drops beneath a set liquidation threshold relative to the outstanding DAI, locked assets may be liquidated. All of this is executed by smart contracts (MakerDAO, 2020). Since smart contracts do not natively know the current value of the US Dollar and relevant crypto assets oracles are needed to provide this information. DeFi has multiplied from roughly 720M USD TVL (Total Value Locked - the value of assets deposited in DeFi protocols) at the beginning of 2019 to a peak of roughly 178B USD at the end of 2021. Currently, TVL hovers between 40B and 45B USD [129]. As mentioned earlier, the rapid growth of this sector has led to many oracle protocols building products targeted specifically to serve DeFi applications.

Another related type of data presently in frequent use is "proof of reserves" data. Stablecoins pegged to currencies, commodities, assets, and their derivatives, are commonly collateralized by assets held off-chain or on another blockchain. In this subcategory, oracles query data sources such as an API made available by a regulated custodian holding funds or assets off-chain or the API of blockchain explorers to verify the amount of funds held by a wallet on another chain [130]. The GBPT stablecoin, for example, is backed by British Pounds (GBP) held in an account at Bank Frick, which provides an API to the nodes of a Chainlink DON. The oracle network pushes new data on-chain every 24 hours or when a deviation of over 2% is reported by the nodes. The oracle network for this PoR (Proof of

 doi:10.20944/preprints202405.0218.v1

Reserves) feed currently consists of 16 nodes operated by independent organizations such as Blocksize Capital, Deutsche Telekom MMS, NorthWest Nodes, among others [131].

Aside from finance, there are many custom types of data being made available through oracles. This subcategory is largely made up of less mature data types that rely on rather primitive oracle implementations. The functionality of many oracle protocols to source data from traditional APIs allows developers to easily source many types of data they may need. Chainlink, for example, calls this functionality "External Adapters" [132], while API3 calls it "Airnode" [96]. While often not optimized for decentralization and trust minimization due to the use of a single data source, it offers accessibility and easy implementation. This category of oracle feed enables some of the most innovative use cases by allowing the integration of sensor and satellite data, among other direct data sources. The example of parametric insurance mentioned previously relies on satellite and weather data to determine whether a claim will be paid out [133]. Another example showing an innovative use case of a direct data feed is the Chainlink hackathon project "Link My Ride" which built an external adapter allowing a smart contract to interface with a Tesla vehicle. This interface was used to offer automated rental functionality, including payment and unlocking the car doors for the renter [134]. While this example was highly experimental, it demonstrates the potential of hybrid smart contracts.

The final subcategory of data feeds is cross-chain interfacing. Here oracles are used to read data on one blockchain and provide that data to a smart contract on another blockchain. Beyond crypto stablecoins, in theory, this allows for blockchain interoperability, such as transferring a token from one chain to another, as discussed in Chapter IV a.

Computation & Functions

The second main category of smart contract inputs provided via oracles is computation and functions. In this category, oracles are used to perform functions that are not natively possible or efficient to do on-chain and provide the output for on-chain use without necessarily using data from any external source [130]. The following is a brief summary of various functions currently in use. While many of these functions have different names depending on the oracle platform being used, we will use the terminology used by Chainlink for simplicity.

Automation

In order for a smart contract to execute a certain function, that function needs to be called. In most user-facing smart contracts, this is of no issue as the smart contract only needs to perform an action when interacted with by a user [130]. This is the case with most DeFi apps, for example, exchanges, which execute a function when a user initiates a transaction. Not all smart contracts, however, function in this way. An example includes a smart contract that is supposed to check a condition every 24 hours [130]. Doing this manually or incentivizing users to do this is not usually efficient or reliable. Oracle networks can be used to trustlessly interact with a smart contract in a specified way. Chainlink's automation, formerly Chainlink Keepers, uses a DON to send a predefined transaction based on a trigger condition. By using a decentralized oracle network, this method offers high reliability and redundancy and incentivizes honest behavior. One example use case for automation mentioned in Chainlink's documentation is a defi yield harvester that withdraws and reallocates capital, generating yield in a defi protocol, allowing for compounding and the employment of advanced strategies [130].

Randomness

Verifiable randomness is currently one of the most prominent use cases for oracles. Computers have two methods to generate randomness: (1) true random number generators (TRNGs) and (2) pseudo-random number generators (PRNGs). PRNGs use a pre-defined algorithm to emulate randomness, whereas TRNGs employ a software-external phenomenon as a source of entropy for randomness generation. Since this external phenomenon cannot be predicted, it may be used to generate a truly random and unpredictable output. In traditional computers, common things used

 doi:10.20944/preprints202405.0218.v1

to generate entropy are, for example, the RPM a certain fan in the computer is spinning, current CPU temperature, mouse movements, etc. [135]. Since Blockchains are composed of thousands of computers contributing to an isolated system, this is not natively possible on most contemporary smart contract platforms. Instead, Verified Randomness Functions (VRF) employ oracle nodes to fill this gap. Chainlink's VRF, for example, works by "combining block data that is still unknown when the request is made with the oracle node's pre-committed private key to generate both a random number and a cryptographic proof" [75]. Since both the randomness and the cryptographic proof are generated together, the output provided on-chain can be checked before making it available for use by a smart contract. Due to this verification step, nodes are unable to provide biased or manipulated outputs, as the cryptographic check would fail. This function is currently primarily used in the generation of NFT attributes, gambling, and blockchain-enabled games [136].

Scalability & Off-Chain Computations

While smart contract platforms are essentially distributed computers, they trade computational performance for decentralization. The trade-off between scalability, decentralization, and security in the context of blockchain networks is known as the scalability trilemma [137]. Certain blockchain platforms maintain a high level of decentralization but are limited in computational ability, while others focus on scalability over security. For computationally intensive operations, it makes sense to move execution to smart contract platforms optimized for scalability or to a secure off-chain environment. Using oracles and cryptographic guarantees, the computation can trustlessly be outsourced from blockchain networks with low scalability and high security to those with high scalability, returning only the result of the computation [138].

Privacy & Security

One common criticism of smart contract platforms is radical transparency, to the point of potential privacy issues [139]. Blockchains are pseudonymous, not anonymous - while it is possible to remain anonymous, patterns and behaviors may be used to identify users, their behaviors, and holdings. This presents a significant issue for institutional users who may need to protect trade secrets, such as the identities of stakeholders in their supply chain [140]. Oracles can be used to deploy privacy-enabling add-ons to smart contracts. While these add-ons are both plentiful and modular, with new ones continuously being built, there are currently three primary applications: zero-knowledge proofs used to confidentially yet verifiably share the outputs of computations (DECO), trusted hardware connections (Town Crier) [141], and the decorrelation of smart contract execution and settlement (e.g. Mixicles) [142].

Zero-knowledge proof integrations via oracles, such as DECO, allow users to verifiably share a piece of data without revealing its full context [132]. For example, a user may prove that their credit score is above a certain number via a direct connection from an oracle node to the credit score provider. The oracle node obtains the relevant number and transmits only it, no other details, in a cryptographically reliable way. This simple mechanic has a wide variety of applicable use cases and could become an integral part of next-generation smart contracts. Examples include age verification, creditworthiness, insurance premium calculations, Sybil resistance, fraud prevention, and more. DECO was purchased by Chainlink from Cornell University in 2020 [143].

Trusted hardware connections allow smart contracts to access temperproof and confidential oracle inputs via just a single node. This means sensitive information, such as passwords, and personal data may be securely stored and made accessible when needed [142]. An example mentioned in "Town Crier: An Authenticated Data Feed for Smart Contracts" by Zhang et al. [141] is parametric on-chain flight insurance. In order to verify and pay out a claim, the smart contract must be able to access the claimant's travel details. If the claimant were to submit his travel details directly to the smart contract, this information would be publicly accessible, infringing on their privacy. Instead, the data could be stored on-chain in an encrypted format. In this case, oracle nodes would be needed to decrypt and

 doi:10.20944/preprints202405.0218.v1

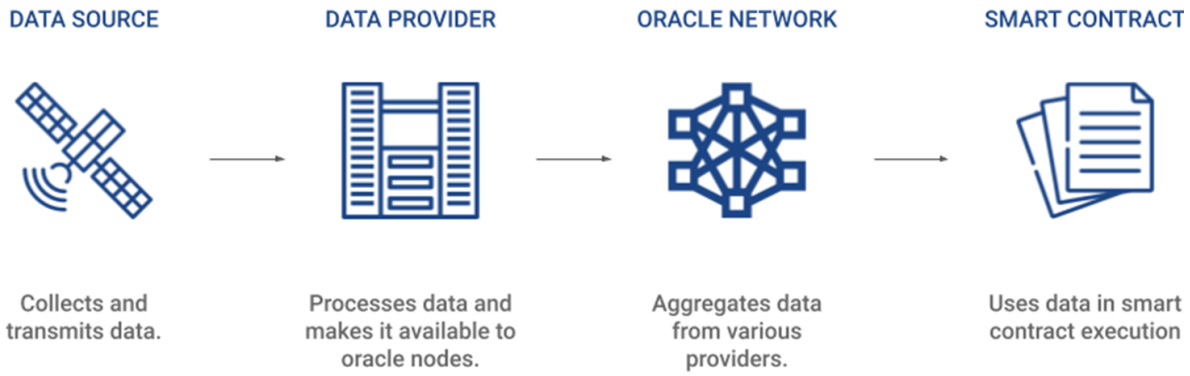
verify the information, potentially allowing the node operators to access the information. Using a trusted execution environment (TEE) solves this. The TEE acts as a black box decrypting and verifying the information and transmitting the result, was the flight delayed or not, to the smart contract [141]. Use cases for this technology may be found in health care, cyber security, and beyond.

Mixicles allow users to detach the outcome of a smart contract from a payment output. Consider a smart contract based on a binary query, e.g., "Was flight X on time?", "Was the price of Apple shares above X at time Y?", "Did Candidate X win the election?". Based on these inputs, an oracle network can tell a smart contract to transfer funds held in a smart contract to any predefined address [55]. By using newly generated addresses, users may conceal the financial outcome of that smart contract, creating a disconnect between the smart contract and their associated transactions [142].

Beyond these functions, there are those built to remedy specific problems appearing across smart contract platforms, such as Fair Sequencing Services (FSS), which protect from front-running exploits in DeFI [144], Off-Chain Reporting (OCF) which enable batching of oracle network responses to save gas costs [130], and many other, at times custom-built, functions. All in all, oracles act not only as secure data adapters but also as a toolbox, adding functionalities to blockchain networks that are not natively possible. Not all oracle implementations support all functions at once. Some may specialize in a single function.

2.6. Stakeholders & Economics

This section explains the stakeholders typically involved, from data sourcing to data delivery across various use cases mentioned in the previous chapter using Chainlink's implementation. The optimal fundamental constellation of stakeholders is a variety of node operators, typically 10-30, sourcing relevant data from a selection of publicly available data feeds [26]. In the case of market data, e.g., cryptocurrency spot trading data, data sources include price aggregators such as coinmarketcap and coingecko, and exchanges such as Binance and Kraken [92]. Data providers for a specific feed are currently curated and allowlisted by Chainlink Labs [93]. Node operators can then choose a data provider to connect to [130]. The network of nodes aggregates this data, and when a trigger condition is met, one node is selected to transmit the data on-chain [26].



Note. Visualization of stakeholders along the oracle data value chain. Own work.

Figure 5. Oracle Data Value Chain.

The above structure pertains primarily to data feeds with sufficient reputable data sources. Newer and more niche data feeds may need to rely on a single data source either due to uncertain demand, e.g., in the case of highly specific satellite imagery, or there simply not being multiple parties who can access and provide specific data, e.g. in the case of bank account balances. Other functions, such as computation, do not require an off-chain data provider, while others, such as Town crier, do not require a network of nodes. In the following, we will go over the major stakeholders in the sourcing and delivery of data on the chain.

 doi:10.20944/preprints202405.0218.v1

2.6.1. Data Providers

As mentioned previously, oracle nodes source data from data providers. Typically, this is done by ingesting an API [130]. This can be a publicly available API or an API made available for this specific use case. While technically, anyone could position themselves as a data provider, there is a particular reputation, transparent methodology, and assurance needed in order to convince node operators that a data feed is trustworthy and reliable. As with Chainlink, some oracle protocols curate a set of data providers for specific feeds. Nodes serving chainlink price feeds, for example, are able to choose from a limited number of exchanges and market data aggregators. Node operators are incentivized to select reliable data providers as incorrect data transmitted via their node may lead to economic or reputational penalties [26]. Hence, it is vital for data providers to have a demonstrably solid methodology for data collection. Considering the example of Chainlink price feeds, there are several options node operators may choose from. Coingecko aggregates market data from over 700 exchanges using a trust score and outlier detection. Coinmarketcap operates quite similarly. Alternatively, exchanges can be used as a primary source for market data, which is reported directly.

Things get interesting when looking at more specialized data types. One example can be found in the case of macroeconomic data. Truflation is a start-up and data provider that provides independent inflation and consumer price index data. Their methodology relies on over 10 million data points to source price data for various individual goods, composing an annually updated household expenditure basket. This allows Truflation to provide close to real-time inflation data [145]. The U.S. Bureau of Labor Statistics publishes inflation data monthly, roughly 12 days after the month's end (U.S. Bureau of Labor Statistics, n.d.). Smart contract developers now have the option to choose between relying on official legacy sources, unofficial sources that update more frequently using their own methodologies, or an amalgamation of both. Ultimately, the user experience may play a large role in developers' decisions. Another interesting example is that of data providers for "Proof of Reserves" (PoR) data feeds. PoR feeds are used to prove the collateralization of certain on-chain assets, such as stablecoins. The usefulness of PoR integration also relies heavily on the methodology employed. In July 2023, an article was published criticizing Chainlink's PoR feeds for being opaque and varying in data source quality [146]. In some cases, these criticisms may be justified as, according to Chainlink's documentation, some PoR data feeds use self-reporting as the data source [130]. Other PoR feeds use a data feed provided by a regulated custodian as the data source, and others use Blockchain Explorer data to source the wallet balance of an address on another blockchain. Not all data providers and methodologies for sourcing data are created equal.

2.6.2. Node Operators

Nodes receive, consolidate, and process data that is either received from off-chain sources or on-chain sources [26]. While anyone can operate an oracle node, it is the case that for more prominent feeds, node operators are curated and allowlisted based on reputation, for example, with regards to Chainlink price feeds. Reputation plays a significant role in getting jobs for a node. Some of the most reputable node operators currently include Blockdaemon, Blocksize Capital, Deutsche Telekom MMS, LinkPool, and NorthWest Nodes, among others [131]. Unlike some other blockchain projects, node operations, in the case of oracles, are often done by specialized institutions [147]. Obstacles for retail node operators include the cost of specific data sources, which may be over $1000 per month to access, and reputation, as smart contract developers would likely prefer a node that has run a certain job a million times with a success rate of over 99,9% when compared to a newer node. Customers may also require nodes for a specific job to deposit collateral in the form of the LINK token to be eligible. In the case of push-based oracles, nodes need to be funded with the relevant smart contracts platform's native currency to pay transaction fees when submitting data on-chain. In the case of pull-based oracles, the requester pays this fee [148].

 doi:10.20944/preprints202405.0218.v1

2.6.3. Customers

Customers of oracle services are the organizations operating the smart contracts sourcing data from oracles. This can be simple smart contracts or complete decentralized applications. One of many examples of a protocol that would not be possible without oracle feeds is Ampleforth. Ampleforth is a cryptocurrency with a targeted market value equivalent to the purchasing power of 1$ in 2019. It uses CPI data and market price data oracles to determine whether Ampleforth is trading above or below the target value depending on which the supply expands or contracts to reach the price target [149]. Oracles allow smart contract developers to not only automate existing business logic but also to implement entirely new logic. As the space is still quite new, it remains to be seen what may be built.

2.6.4. Economics, Incentives, Staking

While some oracle implementations leverage their own cryptocurrency token to incentivize all actors to behave correctly, not all do. This section illustrates how token integration may be used to improve reliability and minimize risk through crypto-economic incentive structures. The Chainlink oracle implementation, for example, uses the LINK token for collateral and fee payment. Node operators put up collateral that is lost if the node acts dishonestly. This allows the network to mitigate the effects of an actor attempting to bribe node operators since the attacker would need to bribe multiple nodes to affect the data output of an oracle network. Meanwhile, the incentives are laid out to exponentially increase the rewards of honesty when other nodes are dishonest. Staking is considered an explicit crypto-economic incentive in addition to which the future fees a node operator stands to earn act as an implicit incentive. This further decreases the incentive for node operators to behave dishonestly since reputational history will impact future earnings [26]. Node operators are considered to be economically rational but not honest by default. Additionally, external actors may also stake LINK to participate in a decentralized alerting mechanism to generate some yield [92]. These stakers are incentivized to send alerts when DONs fail to operate as they should. Cryptoeconomic incentives play a significant role in many oracle protocol designs [150].

3. Literature Review

This section summarizes opportunities, issues, and discussions found in academic research to identify research gaps. First, it goes over the current state of research regarding blockchains and smart contracts in general. Second, it summarizes open questions and identifies challenges regarding the implementation of oracles.

3.1. Summary of Research

While the topic of oracles is still quite niche and largely unknown outside of the blockchain field, numerous papers have been published. These papers largely fall into three categories: (1) hybrid smart contracts, (2) oracle technology and implementations (3) specific use cases. In the context of this paper, (2) and, to a certain extent, (3) are the most relevant.

The 2016 paper "The Blockchain as a Software Connector" by Xu et al. is among the first independent research papers discussing off-chain and on-chain communication and data transfer in the context of smart contracts. A key argument made in this paper is that in certain transactions, first-party oracles are sufficient. The example mentioned involves a person renewing their driving license. In this case, it is assumed that the person already trusts the issuing party, and there is no need to decentralize the data feed or employ a third party. If there is no trust between the transacting parties, a "validation oracle" may be needed.

In 2017, Chainlink released its initial whitepaper [151], just 14 days before its initial token sale (ICO) De Collibus et al. [152]. The whitepaper was a complete, yet purely theoretical, introduction to Chainlink's proposed solution to the oracle problem. Aside from the initial design, the paper highlights a need for composability - building simple tools that can be combined to form complex systems. While this paper mentions the concept of off-chain data aggregation, it falls short of mentioning DONs.

 doi:10.20944/preprints202405.0218.v1

Several of the decentralized computation functions mentioned in the theory section are only mentioned briefly or not discussed. Nonetheless, the release of this whitepaper represents a significant milestone for, in some ways, the birth of the decentralized blockchain oracle space.

In the paper "The Oracle Problem - An Analysis of How Blockchain Oracles Undermine the Advantages of Decentralized Ledger Systems", Egberts [153] discusses how centralized oracles pose a single point of failure and methods such as multi-source data feeds and reputation systems that may help solve this problem. The influence of Egbert's writing is evident through the way it inspires other papers, such as Breidenbach et al. [26]. Eberhardt and Tai [154] published a paper highlighting the need for a way in which blockchains may be able to move data off-chain for computation. While not explicitly mentioning oracles, the paper concludes, "We still consider off-chaining techniques to be key tools in blockchain-based application engineering as they introduce additional functionality and potentially significant cost benefits" [154]. As discussed previously, oracles now offer a variety of off-chain computation tools.

2018 and 2019 saw the launch of the first wave of dApps and a variety of other novel blockchain applications. Some of the core learnings about practical blockchain application design gained during this time are synthesized in the book "Architecture for Blockchain Applications" [155]. Including the three ways in which blockchains may communicate with the real world: oracles, reverse oracles, and pairings of legal and smart contracts (p. 113). According to the book, oracles have two main drawbacks. First, trust - oracles vary in quality depending on implementation from centralized first-party oracles to decentralized oracle networks. Users of Oracles must understand this and select an implementation that respects the given circumstances. Secondly, validity - data can not natively be verified by the consensus of a given chain; there is a full reliance on the oracle (p. 117).

By 2020, the oracle problem began receiving more attention, with several directly related papers being published. Giulio Caldarelli published a systematic literature review (2020a), which found that in most papers addressing a use case for which off-chain connectivity is needed, the oracle problem was not taken into consideration. Shortly thereafter, he published a journal article titled "Understanding the Blockchain Oracle Problem: A Call for Action" (2020b). This article hypothesizes oracle implementations for various use cases and raises questions about the reputation of data sources: Can a firm's reputation alone counter the oracle problem in supply chain use cases? Can patients in a health care system act as oracles? Is it at all possible to manage an energy market platform without a central authority?

2020 also marked the publishing of one of the most commonly cited papers in the oracle field, "Trustworthy Blockchain Oracles: Review, Comparison, and Open Research Challenges" by Albreiki et al. (2020). This paper compares the approaches of a variety of decentralized oracle protocols, including Augur, Chainlink, Witnet, ASTRAEA, and Aeternity. Challenges identified include developing ways in which smart contracts may feature integrated fail-safes in the case of unintended data and implementing ways to cryptographically verify data received from oracle nodes. Such methods are now being implemented by various projects, but it is unclear whether it has become a widespread industry standard. Most notably it concerns the trustworthiness and decentralization of actors participating in oracle networks [156]. The importance of decentralization with regard to the reliability of oracles is also highlighted in the paper "Reliability Analysis for Blockchain Oracles" [157] and the paper "Blockchain Oracles: A Framework for Blockchain-Based Applications" (Memmadzada et al.,2020) which explain how centralized oracles may work well for permissioned systems, but multi-source oracles are better suited for situations with a large number of actors. Also noteworthy, in 2020, Towncrier had not yet been purchased by Chainlink and was mentioned by several papers as a standalone oracle implementation.

In December 2020, the World Economic Forum published a whitepaper titled "Bridging the Governance Gap: Interoperability for blockchain and legacy systems" written by Chainlink Labs in collaboration with the Blockchain and Digital Assets team of the World Economic Forum [158]. The report gives an example of how India's national crop insurance portal (NCIP) could effectively

 doi:10.20944/preprints202405.0218.v1

leverage decentralized oracle feeds to validate data. Going on to state, "An oracle network with a proven incentive and reputation system would also unlock a marketplace for the larger open source community to enrich NCIP's forecast systems with increasingly localized and granular data." In this case, oracles would be able to augment a legacy system to improve the reliability of weather information systems. Other examples mentioned include customs processing and vehicle registration systems. These cases demonstrate how smart contracts may be integrated with current infrastructure rather than disrupting them outright.

Another relevant 2020 paper, "Blockchain for COVID-19: Review, Opportunities, and a Trusted Tracking System" (Marbouh, 2020) proposes a dApp to track new cases, recoveries, and deaths due to COVID-19. The proposed application uses multiple trusted oracles to get relevant data from various official sources. The oracles send the data to a smart contract that registers the data and which oracle broadcast it. The data is aggregated, and oracles are assigned a trust score based on how closely their data matches the aggregated values. While using multiple oracles does "avoid the occurrence of a single point of failure," as stated in the paper, this implementation does not mention any economic incentives for individual nodes to behave honestly. This is likely not relevant since nodes are considered trusted. If nodes were untrusted, i.e., with anyone able to join as a node operator, this implementation would be insufficient. The paper does highlight ways for permissioned systems to natively integrate oracles via smart contracts.

2021 saw Giulio Caldarelli continue publishing research about blockchain oracles, focusing on separate use cases such as the fashion supply chain [72], decentralized finance [159], and wrapped tokens [88]. The paper "Verifiable Computing Applications in Blockchain" [160] briefly mentions the use case of verifiable randomness (VRF) in the context of verifiable computation. Another paper from this year [127] closely examines Chainlink oracle activity, providing insights regarding the demand for data, node earnings, and data regarding data feeds. According to this paper, Chainlinks customers were primarily interested in price feeds for DeFi applications, with 75% of price feed traffic originating from the DeFi project Synthetix. Interestingly, at the time of writing, Synthetix uses Pyth in addition to Chainlink. Furthermore, the paper speculates that high fees on smart contract platforms at the time may be a reason the demand for oracle feeds outside of price and market data was low. The paper concludes, "...the Chainlink ecosystem on the Ethereum network appears to be driven purely by DeFi's demand for decentralized market price feeds".

A Systemization of Knowledge titled "SoK: Oracles from the Ground Truth to Market Manipulation" by Eskandari et al. [161] provides a thorough summary of available research at the time. Included in the paper is a table comparing available oracle implementations and modules. Attributes mentioned include how reporting nodes are selected from a decentralized network, types of data sources (categorized into API, Human, Smart Contract, and HTTPS), aggregation mechanisms, and more. Considered are the possibly largest number of oracle projects up to this point. Most projects mentioned in this SoK are still active today and are also considered in this paper. Several important considerations regarding the crypto-economic systems supporting oracle networks are also discussed in this publication. One such insight is that on-chain modules on public blockchains may lead to high fees for data consumers, potentially disqualifying applications with a need for frequently updated data and limited proportional revenue. This may explain why DeFi applications make up the largest portion of oracle data consumers; here, data is used to directly facilitate revenue-generating activity. The paper concludes with the warning that the failure of a prominent oracle project may lead to a chain reaction with applications using this specific oracle implementation failing as well. While several DeFi applications today integrate a secondary fall-back oracle it would be important to see how commonly this is.

2021 also marked the publication of Chainlink's updated whitepaper "Chainlink 2.0: Next Steps in the Evolution of Decentralized Oracle Networks" [26]. This paper was another major milestone as the previous whitepaper, acting as the primary document explaining how the network works, had been published in 2017, two years before Chainlink went live on mainnet. The years of practically working

 doi:10.20944/preprints202405.0218.v1

and experimenting with a live network have allowed Chainlink labs to iterate on and expand the functionalities of its oracle implementation. For comparison, the initial whitepaper is 38 pages, while the whitepaper 2.0 is 136. Three authors contributed to the initial whitepaper, while the whitepaper 2.0 credits 14 authors. The Whitepaper 2.0 explains in exact detail how Chainlink's oracle implementation operates regarding security, crypto-economic incentives, and how specific functions contribute value. All in all, there is a remarkable difference in research attention dedicated to oracles between 2017 and 2021. Of course, in 2017, oracles were in their infancy, and even today, the topic is still relatively unknown in the great scope of things.

In 2022 a number of relevant papers were published. "On the Integration of Blockchain With IoT and the Role of Oracle in the Combined System: The Full Picture" by Sadawi et al. (2022) discusses the importance of oracles when integrating IoT with blockchain, something that was not extensively researched previously. In the scope of the paper, the authors set up a prototype CO2 sensor able to communicate with Ethereum smart contracts. While this was only a basic prototype using a centralized, first-party connection, it helps illustrate the means by which sensors may make data available on-chain to those unfamiliar with the topic. Pasdar et al., meanwhile, published a paper focused on the technical underpinnings of various oracle implementations (2023). The oracle protocol selection taken into account in this paper is based on Egbert's 2017 paper. Pasadar et al. argue that oracle implementations largely fall into one of two categories: voting-based oracles and reputation-based oracles. Voting-based oracles are built centered around explicit incentives, while reputation-based oracles are focused on implicit incentives.

The paper "Toward Trustworthy DeFi Oracles: Past, Present, and Future" [144] compares the performance and trustworthiness of popular oracle implementations, including Chainlink, Band, DOS, Nest, and Witnet. Furthermore, the paper outlines a rubric by which the trustworthiness of an oracle may be evaluated. Namely, accuracy - as in accounting for bias from data, time-efficiency, scalability, and security - short-term and long-term adversarial costs. The paper also points out that major DeFi protocols may begin operating as oracles themselves, providing market data feeds natively and directly.

2022 also saw the publication of two papers addressing oracles and associated research directly: "Blockchain Oracles: State-of-the-Art and Research Directions" [162] and "Overview of Blockchain Oracle Research" (Caldarelli, 2022). The former highlights several open challenges faced by oracle implementations as of the time of writing. Mostly, these have to do with a need for deeper evaluations concerning cost, performance, and security. This is an important point, as some oracle implementations of crypto-economic systems are incomplete or subsidized with the likely goal of increasing adoption. The latter paper evaluates the amount of research done relating to oracles in the seven years prior to its publication, finding a total of 162 relevant papers. It concludes that blockchain oracles are "still a widely neglected subject" despite their crucial importance in securing decentralized applications. This is likely to change in coming years as oracles find more use cases and implementations are iterated upon.

Now, in 2023, research continues. In the paper "Before Ethereum. The Origin and Evolution of Blockchain Oracles" Caldarelli and Ellul [72] shines a spotlight on the beginnings of the oracle field. In 2011, Satoshi Nakamoto mentioned the possibility of running scripts dependent on external data on Bitcoin. Developer Mike Hearn recognized potential issues with relying on a single data source and theorized how an oracle implementation may overcome this. As a result, there were several little-known attempts to implement trust-minimized oracles on Bitcoin, including Oraclize, ORISI, Reality Keys, Counterparty, and Truthcoin. Ultimately, oracles were never widely adopted on Bitcoin, likely because smart contracts were far more easily implemented on purpose-built blockchains, and many Bitcoiners were opposed to storing external data on-chain. This paper shows that a small number of developers have long been familiar with and attempting to solve the oracle problem. Unfortunately, most early Bitcoin developers primarily shared insights via private interactions, email chains, and forum posts, not academic publications.

 doi:10.20944/preprints202405.0218.v1

Compared to the early days of trying to solve the oracle problem on Bitcoin, there is now a significantly higher amount of academic interest in the topic, and there is likely even more developer interest. While Caldarelli is right in pointing out that this topic is still widely neglected, considering its importance, this may be changing.

3.2. Open Questions and Challenges

Considering the aforementioned literature, there are clearly still plenty of little or completely unexplored topics, open challenges, and unanswered questions. Largely, research impulses regarding blockchain oracles are twofold. On the one hand, explicitly defined questions outlined in the existing literature. On the other hand, implicit research impulses based on little-explored challenges and untouched topics. This chapter explores these research impulses with the goal of synthesizing specific research questions, which may later serve as a basis when defining interview questions to be tailored to industry insiders and researchers.

Firstly, as previously discussed, the book "Architecture for Blockchain Applications" [155] outlines two defining factors that are vital when building trustless hybrid smart contracts: trust and validity. Trust refers to the structure of a given oracle implementation. The degree of decentralization applied with regard to a specific oracle structure, from data creation to smart contract execution, plays a decisive role in both the functionality and security of a smart contract. In a proof of reserves use case, for example, it may only be possible to have one data provider, which prompts the question of whether a decentralized network of nodes adds an additional level of reliability. In other cases, such as price feeds, data aggregation does improve reliability. A question that arises here is whether a specific oracle implementation is superior to others between networks of off-chain decentralized oracle nodes, a separate blockchain as part of the oracle implementation, or aggregating data on-chain after receiving it from a variety of first-party oracles. It appears as though the ideal implementation may be case-specific. Looking further into this point would also address the question raised by Caldarelli et al. [23] regarding whether a firm's reputation alone might be enough to counter the oracle problem in supply chain use cases. A comprehensive comparison of different use cases and available solutions would be helpful for smart contract architects.

"Demystifying Pythia: A Survey of ChainLink Oracles Usage on Ethereum" [127] points out that "the number of individual users of the Chainlink platform is not very high." and that Chainlink price feeds were the most in-demand. A few other publications mentioned above seem to confirm this sentiment. Meanwhile, several newer oracle platforms are focused exclusively on price feeds. This raises questions about other use cases, such as supply chain, healthcare, etc., which have been explored in research. Overall, the growth and discovery of practical hybrid smart contract use cases seem to be one of the greatest challenges to them being widely implemented. This raises several questions about the broader utilization of oracles beyond market data. For instance, if such utilization is low, could this be attributed to particular factors like economic constraints or challenges in identifying and developing distinct use cases? Additionally, it is worth considering whether the demand for specific data feeds influences their supply or vice versa. Lastly, whether the future will bring a single, universal oracle solution or multiple specialized implementations catering to varied use cases, such as price feeds, randomness, and sensor data, remains to be seen.

Another general question brought up is regarding the maturity of the oracle space. Several papers mention the state of the oracle space at the time of writing. Has there continued to be rapid progress as there was between 2017 and 2021? What future developments are on the horizon? Understanding where the oracle space is currently and how experts see it developing both in the near and far future could provide an updated outlook regarding greater trends in the hybrid smart contract and, to an extent, the DeFi, space.

As discussed in the first part of the literature review, the publication "SoK: Oracles from the Ground Truth to Market Manipulation" [161] brings up two important points. Firstly it highlights two rarely discussed risks: the potential fallout for external projects in the unlikely case an oracle project or

 doi:10.20944/preprints202405.0218.v1

implementation fails and that the economic security of token-based oracles may be undermined in case the token drastically loses value. The former case could be addressed by using fall-back oracles or possibly by adding another level of aggregation. The area of integrated fail safes and data verification was also mentioned as an open research area in "Trustworthy Blockchain Oracles: Review, Comparison, and Open Research Challenges" [156]. "Blockchain Oracles: State-of-the-Art and Research Directions" [162] goes on to state the need for deeper evaluations concerning cost, performance, and security.

Furthermore, Eskandari et al. [161] point out that oracles may employ entirely different methods of sourcing data. While some use an API, others may gather data directly from people, among other methods. As discussed earlier, selecting the ideal constellation from data sourcing to data delivery depends heavily on use cases, with not every case being as decentralizable as others. In some cases, there is only one original source of the data, for example, when asking a patient how they feel. There is no way to decentralize that data anymore, and adding any further intermediaries may introduce points of failure. This raises the question of how difficult it is to develop a proper methodology for data provision for a given use case, even within the framework of existing oracle implementations. This also links back to the earlier questions regarding the general development process of new use cases. Furthermore, it might be possible to develop a rubric by which new use cases and their difficulty of implementation may be evaluated.

Topics that seem to be neglected in research are questions regarding the relationships between the various stakeholders involved in providing data feeds to smart contracts outside of purely economic aspects and opportunities for coaxial business models that may either build on present constellations or create value by augmenting and improving them. Additionally, it may be interesting to ask different stakeholders directly what use cases they are most excited about to better understand what drives and motivates their involvement in such a new industry.

3.3. Focus of this Paper

Combining the initial research impulses laid out at the beginning of this paper with open and unexplored questions identified above, this paper will set out to contribute to the current research by providing insights regarding the following general questions. The literature review points to three guiding research questions. (1) How do stakeholders position themselves within the competitive landscape, and what are the criticisms and advantages of their approaches? (2) How mature is the industry in its current usage, and what is its future outlook? (3) What opportunities are enabled by the industry, and what tangential businesses may arise from it? The specific list of questions that will be used to guide unstructured interviews to be conducted hereafter is discussed in the methodology section. What follows is an overview of the research conducted in the scope of this paper. On a high level, the research areas to be explored can be broken down into three main categories:

a) Industry Landscape

   This includes questions regarding current stakeholders in the data delivery process, their interactions, and incentives. Furthermore, speaking with experts developing different oracle implementations may provide insights regarding the competitiveness and positioning of oracle projects. Also, we will explore the overall industry maturity and future outlook as perceived by these experts.

b) Innovation & Opportunities

   Here, we explore stand-out use cases beyond what exists today and opportunities for auxiliary business models that may improve processes along the value chain from data sourcing to data delivery. Finally, this category explores the process by which new use cases and in-demand data types are identified, implemented, and ultimately scaled.

c) Challenges & Implementations

   Focuses on elements of decentralization, security, validity, and reliability. Additionally, which data types are readily available and in demand, and which data types are difficult to make

 doi:10.20944/preprints202405.0218.v1

available on-chain? Furthermore, different design elements and the positioning of various oracle implementations to overcome challenges will be explored.

The primary goal of this paper is to contribute to the current research by evaluating interviews with a wide variety of industry stakeholders, including oracle projects, node operators, and data providers, along with related companies and researchers. More details about the interviews and how collected data is processed are described below.

4. Methodology

This section explains how data was collected and evaluated, from creating a list of guiding interview questions to transcribing and deriving an assessment of the insights highlighted. The purpose of this chapter is to outline the chosen process and methods. Considering the three high-level topics to be investigated, described in detail above, Industry Landscape, Innovation & Opportunities, and Challenges & Implementations, and the existing literature on the topic in question, a qualitative research approach was chosen.

The exact methodology selected for data collection for this paper took the form of unstructured interviews with experts working across the oracle industry. These interviews were conducted remotely via a video call. The interview audio was recorded and transcribed via a reputable automated transcription service. Initially, a pool of roughly 60 potential interviewees was generated, including the teams behind the leading oracle protocols, large node operators, data providers, other types of companies involved with the oracle space, and independent researchers. Ultimately, a total of 15 such interviews were conducted, each ranging between 30 and 50 minutes.

The recorded interviews were transcribed using a digital transcription tool. Subsequently, the transcriptions were analyzed to highlight recurring themes, divergent views, and points of interest. As previously stated, the interviews were largely unstructured. This enabled specific follow-up questioning based on individual background, expertise, and interests. Questions used to guide the conversation included “What attracted you to work in the oracle space?”, “What types of data that would be the most useful to smart contract developers are currently missing?”, and “Could you name a specific use case for oracles that you find particularly exciting?”. The questions were aimed at inspiring conversation providing insights about the three high-level topics mentioned above.

All transcripts were reviewed for statements and insights regarding the three categories of exploration: Industry Landscape, Innovation & Opportunities, and Challenges & Implementations. Matching statements were marked, compiled, and evaluated according to these categories and the sub-questions pertaining to each topic. The approach for evaluating interview content is an exploratory, inducive one with the goal of contributing qualitatively to the primarily quantitative research published in this area. Transcripts of the individual interviews can be found in appendix B.

5. Results & Discussion

In this chapter, we delve into the outcomes of the 15 interviews conducted with insiders familiar with the current landscape of oracle implementations and stakeholders. First, the participants and their backgrounds will be discussed in depth. Second, this section highlights specific insights addressing the three main research areas: Industry Landscape, Innovation & Opportunities, and Challenges & Implementations. The purpose of this section is to summarize and discuss key insights in these four areas.

5.1. Interviewee Pool

The final pool of interviewees comprised six interviewees working at or representing oracle protocols, four interviewees are professional oracle node operators, two represent data providers. Additionally, two of the interviewees are independent researchers specializing in the oracle space, and the final interviewee represents a start-up leveraging oracles in their business model.

 doi:10.20944/preprints202405.0218.v1

The following six oracle protocols are represented among the interviewees, who all work in developmental or leadership positions within these projects: API3, Pyth, Redstone, Tellor, Witnet, and Supra Oracles. Among the four node operators included in this study are Blocksize Capital, Blockdaemon, and Northwest Nodes. One of the two data providers, Truflation, sources CPI data and provides an inflation data feed for the US and Great Britain. The other operates a proof of reserves data feed for a fiat-collateralized stablecoin. One of the researchers has contributed significantly to the available academic literature about oracles, while the other has acted in an advisory role during the early stages of two separate oracle protocols - Chainlink and API3. Finally, the start-up included in this study Hyphen.Global employs sensor and satellite data to provide verified data streams related to climate and emissions data.

Ultimately, this composition of experts working within different areas of the oracle space with individual motivations and experiences should be able to provide holistic insights into the questions described in the previous chapter. Since there is little relevance as to which speaker expressed which opinions, the following sections will only refer to "interviewees 1-15" rather than using names or other identifiable information unless strictly significant.

5.2. Industry Landscape

This section synthesizes statements from the 15 interviews, focusing primarily on current stakeholders like oracle projects, data providers, and node operators. While some discussion of individual oracle implementations is included, a comprehensive exploration of decentralization is reserved for Part IV of this chapter. Here, we delve into issues such as industry competitiveness, the specialization of various implementations, and overall industry maturity.

5.2.1. Market Dominance, Competition, and Implementation Design

Firstly, a significant number of interviewees pointed out Chainlink's role as the current market leader. This sentiment was echoed, at least to some extent, by interviewees 2, 3, 6, 7, 11, 12, and 14. Interviewee #2 points out that an estimated 80-90% of revenue generated for data providers and node operators was on the Chainlink network. Interviewee #6 stated, "We can agree that Chainlink is the market leader in terms of integrations and functionality," while also highlighting that other oracle implementations had the same opportunities Chainlink did but failed to capitalize on them. Interviewee #14 points out that Chainlink's current market position is attributable mainly to first-mover advantages and significant funding raised early on. Interviewees #3 and #11 also pointed out the foundational role Chainlink has been playing in the context of DeFi.

While Chainlink currently holds a dominant position in the oracle market, several interviewees pointed out perceived design flaws and criticisms with this implementation, to be discussed in section IV. Correspondingly, most interviewees agreed that the market for oracle implementations was not a "winner take all" one, with different designs being more optimal for certain use cases. Interviewee #3 expressed the concern that while unlikely, there is a potential for the industry to evolve towards an oligopoly with a few very large players controlling the significant majority of the market, similar to Apple and Google's dominance over app markets.

While certain oracle protocols aim to offer a full suite of services, others have specialized in various niches, allowing for the implementation of specific elements optimizing oracles for exactly those use cases. While there will be a more complete look at different implementations and their trade-offs in section IV, this is an overview of the considerations highlighted in the interviews.

Interviewee #1 highlighted the difference between push- and pull-based oracle design. Push oracles actively send data to the blockchain, while pull oracles are ones where users or contracts request the data they need. These contrasting implementations vary in the frequency of updates and in the cost of transmitting those updates on-chain. In the case of push-based oracles, the transaction fee associated with data transmission is paid by the oracle node operator. The data can then be accessed by a variety of protocols. However, the protocols rely on the node's update frequency and may not

 doi:10.20944/preprints202405.0218.v1

have access to current data when needed. In the case of pull-based oracles, updates are requested by a smart contract or dApp, with the respective operator paying the transaction fees to make the data available on-chain. While this relieves node operators and allows dApps access to data the second it is needed, depending on the smart contract platform, this may, in some cases, be prohibitively expensive. Interviewee #1 foresees a future where there will be increased emphasis on hybrid approaches, with there being a general shift towards pull-based oracle designs.

The cost of providing data on-chain was highlighted as playing a significant role in the future design of oracle implementations by Interviewees 1, 3, 4, 7, 8, 11, and 15. In addition to push vs. pull-based oracle design, some implementations have decided to utilize a purpose-built blockchain, among other fee optimization methods.

Furthermore, oracle design varies based on the type of data being transmitted. According to interviewee #15, there are generally three main designs. (1) Algorithmic oracles - Suitable when multiple reliable data sources exist, e.g., price feeds and market data. (2) Optimistic oracles - Used for specific data where there might not be multiple authoritative sources, and the data is only needed once. (3) Single source design: Risky as data comes from only one authoritative source, which could be faulty, e.g., bank account balances or vehicle odometer miles. More on the topic of fee optimization and challenges with different data types to be discussed in section IV.

Different implementations enable different use cases. As projects experiment with different designs, use cases are likely to develop around them, with newer entrants being able to evaluate market needs before designing their implementations. This theory supports a future scenario with many different oracle projects working in different niches rather than there ultimately being one dominant project offering a suite of products that effectively addresses all the various needs of different projects. The interviewee states, “Specialized solutions are gonna win out [in the long run]”.

5.2.2. Maturity & Current Use Cases

Interviewee #14 estimates that, outside of experimental use cases, oracles today are used almost exclusively in the context of DeFi, mostly to provide data necessary for accurate collateralization requirements in Lending protocols. The interviewee went on to state that DeFi has been a huge driver for the oracle space as it could arguably be “the only thing that is actually working inside the smart contract space” - in the context of decentralized applications generating enough revenue to be able to afford oracle data. Similarly, the interviewee argues that the “NFT craze a couple of years ago” led to a significant amount of demand for verifiable randomness. While these two use cases, price feeds and randomness, currently make up most of the demand for oracles, there are clear intentions to move beyond this.

The company one of the interviewees formerly worked at operated as both a data provider and a chainlink node operator, respectively. Considering their already existing business model, they had a large quantity of reliable market data available and decided to monetize this via oracles. According to this interviewee, setting up an oracle node is relatively easy for a team of DevOps specialists. The only significant barrier node operators may face is the cost of data they forward, which needs to be paid to data providers.

The company interviewee #9 is working at has been acting as a data provider for several years now. This interviewee highlighted that there is no significant effort or expense associated with providing PoR data to a Chainlink DON. On the other hand, in the case of this business, data provision is seen as experimental and not a significant source of revenue. In this case, the company only began providing data as an auxiliary service for an existing customer. While the above examples highlight the fact that traditional companies could easily enter the market for Web3 data provision, there currently is little financial incentive to do so.

On the other side of the calculation, there must be projects willing to pay for oracle data in order to sustain the ecosystem. While there may be few barriers to entry for data providers and node operators, aside from the cost for data in the case of node operators, the cost may be more of a barrier in the case

 doi:10.20944/preprints202405.0218.v1

of data consumers. Interviewee #11 highlights that business models need to be sustainable in order to pay for oracle services, and not every project has figured out how to become sustainable. Interviewees 3, 4, 6, 7, 10, and 13 mentioned how in the past, Chainlink subsidized node operations. Since Chainlink oracles are primarily push-based, which resulted in, as Interviewees #10 and #13 put it, a situation where Chainlink oracle data was basically free for protocols. Now that subsidies are winding down and data users are expected to pay for data used, there may soon be a pivotal moment in which we will see the actual paid demand for oracle services.

According to interviewee #11, an executive officer at Redstone, there is little doubt about their ability to attract paying customers. "When we talk with projects they are willing to pay..., [they say:] I mean like we understand we need to pay it is just a matter of making the price reasonable not like through the roof". Nonetheless, use cases for oracle feeds currently seem to be largely decided based on the project's ability to generate enough revenue to pay. It will be important to reduce prices in order to allow for more far-reaching adoption.

While oracles are agreed upon to have come a long way since the initial designs, there is still a way to go in order to fully solve the oracle problem for all use cases. Additional challenges are highlighted in section IV.

5.2.3. Future Outlook & Stakeholders

According to interviewee #8, there is an expectation that many projects and features that have been in the design and testing phases for several years will launch in the coming 1-2 years. Within 5-10 years, interviewee #8 foresees hybrid smart contracts and oracles becoming mass-market ready. The interviewee envisions a smooth transition from legacy systems to hybrid smart contract ones, stating, "Most users won't notice the difference as these technologies will operate in the background, providing improved efficiency and security.'

This merging of legacy systems and hybrid smart contract-based systems may come in the form of bringing real-world assets on-chain, either via tokenization or collateralized representation. Interviewee #9, who has a background in the legacy financial system, notes that tokenized securities drastically increase efficiency as they could eliminate the central security depository, reduce costs, lower settlement time, and offer transparency. The significance of tokenization cannot be understated. Interviewee #2 believes that the tokenization market is going to be several tens of trillions of dollars in volume and that since institutions will need to provide data proving ownership of the tokenized asset, oracles will be of significance. Interviewee #3 underscores this:

> [there] might be a big market for it [tokenization] because obviously [the] financial market and financial industry is so big, you know, ten times or a hundred times bigger than I ever thought in my wildest dreams, even having studied economics, but it's just so, so gigantic.

While tokenization may be where the true potential of blockchain oracles lies, this use case requires there to be more legal clarity. In the meantime, Interviewee #2 sees much potential in perpetual futures. A perpetual future theoretically allows users to take long or short positions on any asset or event through a data feed and smart contracts. Interviewee #1 gives examples: "You can long the annual radiation from the sun, or you can short the amount of traffic on a road". These contracts could have interesting applications, such as automated hedging systems by linking interconnected perpetuals. Interviewee #8 highlights legal uncertainty and regional restrictions regarding perpetual futures. However, as perpetual futures are more permissionless than tokenized assets, they may remain popular until tokenized assets are widely available.

Interestingly, the current oracle industry seems to be a rather small field, with many stakeholders we spoke to mentioning one another in various contexts. Due to the immature nature of the oracle market, stakeholders need to collaborate and communicate in order to implement custom data feeds. There appears to be direct communication between the various data users and node operators. Node operators and the teams building various oracle implementations are also commonly actively communicating.

 doi:10.20944/preprints202405.0218.v1

5.3. Innovation & Opportunities

This section presents insights gained throughout the interviews regarding use cases for hybrid smart contracts beyond what exists today and opportunities for auxiliary business models that may improve processes along the value chain from data sourcing to data delivery. Finally, this section will attempt to illustrate the process by which new use cases and in-demand types of data are identified, implemented, and ultimately scaled.

5.3.1. Business Models

Interviewee #1 highlights that there are already a large number of companies whose business model revolves primarily around sourcing, analyzing, and selling data. While selling this data to on-chain users via oracles is not the main focus for a vast majority of such companies, this may change. We spoke earlier about the ease with which companies may enter the data provider market. However, due to the relatively small amount of revenue currently generated through oracles, many large legacy institutions, such as financial services providers, auditing firms, and similar companies, do not currently appear to be moving in this direction.

On the other hand, there are companies with oracle data provision as a core part of their business model. Two examples of such start-ups are Hyphen.Global and Truflation. While these companies both also sell their data to Web2 customers, they appear to primarily position themselves as a foundational infrastructure provider for on-chain applications. Based on our interview with the CEO of Hyphen.Global, it was explained that they are positioning themselves to verify the quality of on-chain carbon credits. Similarly, in our interview with the founder of Truflation, they expressed excitement about the potential of algorithmic stablecoins tied, for example, to the cost of living. Considering the potential of both tokenized carbon credits and inflation-resistant currencies, this positioning to capture future market share appears very promising. Nonetheless, there do not yet seem to be too many comparable companies.

Overall, identifying and implementing potential use cases for hybrid smart contracts poses a complex challenge. The dilemma arises not only from the need to build the smart contracts themselves but also to establish the accompanying data feed infrastructure. A key question is whether new data feeds should be developed to inspire fresh use cases or whether a specific demand for certain data types must precede such development. This conundrum was explored in Interviews 2, 3, and 7 and likened to the "chicken or egg" scenario. Interviewee #7 emphasized that the quandary of whether to develop products without guaranteed user demand is a major business challenge, not just in the oracle arena but in the broader crypto space as well. Interviewee #3 noted Chainlink's strategy of subsidizing node operations as a solution to this challenge.

Interviewee #2 disagrees with the pretext of a chicken-or-egg problem, pointing out that bringing data on-chain via adapters or API connections is very easy. The challenge lies in achieving optimal decentralization for the use case, taking into account the amount of value secured. This concept of projects moving from centralization to decentralization as they mature was discussed in various interviews and termed "progressive decentralization".

Interviewee #7 brings up a point about projects relying on each other to integrate. Projects build infrastructure, and that infrastructure needs to be used and paid for. Data connections via oracles operate as micro-services, which can be combined to generate value. Interestingly, an example of this type of combinatorial innovation can be seen within the circle of our interviewees as Hyphen.Global contributes climate data to Truflation's CPI metrics. All in all, though, estimating the amount of use and potential revenue generated by a given opportunity is extremely difficult, and new projects may need to be subsidized. Related to the problem outlined above, interviewee #3 highlighted that it does not matter how reliable and secure your hybrid smart contract implementation is; end users need to trust it; otherwise, they will not use it. The success of hybrid smart contracts relies on factors like track record, word of mouth, and social trust. According to interviewee #3, the true challenge for implementing new

 doi:10.20944/preprints202405.0218.v1

use cases is raising awareness among potential beneficiaries. A potential opportunity for auxiliary businesses that help bridge this gap was suggested.

Overall, outside of DeFi, use cases for data feeds are still relatively undeveloped. There are many experimental implementations, many working together, though it remains to be seen whether the demand for hybrid smart contract-based applications will be as high as anticipated by these actors. Some issues still need to be overcome, as will be explored in section IV.

5.3.2. Innovation

This section highlights the interviewee's statements related to specific use cases that appear especially promising or personally interesting. As discussed earlier, the integration of real-world assets emerged as the general theme of future use cases, showing great potential during a majority of interviews, namely 1, 2, 3, 4, 7, 8, 9, and 15. The integration of real-world assets into hybrid smart contracts will significantly increase the value they are able to offer. This section will discuss four specific use cases for oracle feeds and hybrid smart contracts. Perpetual futures, supply chain tracking, carbon certificates, and asset-backed currencies

Interviewee #1 highlights oracles may allow for the integration of real-world assets without requiring a mature legal tokenization framework, stating, "you don't need to tokenize every asset, like there isn't really an advantage to doing so, to make a derivative out of literally anything in the world, all you need is an X over Y data feed". On the other hand, derivatives are naturally not collateralized 1:1 by the asset they represent, potentially leading to issues in certain scenarios. Furthermore, according to interviewee #8, these types of futures are banned in certain jurisdictions. Looking past perpetual futures, the general process of representing an asset on-chain brings with it case-dependent opportunities and challenges. During our interview process, there was specific mention of carbon certificates, securities and financial instruments, real estate, and other derivative instruments tied to real-world assets and economics.

Interviewee #9 brings up the idea of tokenizing shipping containers while they are in transit between ports in order to utilize them as collateral for new credit lines. Interestingly, the supply chain area of application was also brought up organically in interviews 3, 6, and 15. Interviewee #3 is critical about the application of oracles to overcome issues with on-chain supply chain tracking saying "oracles might be able to help in some sense." However the issue at large is not data immutability and reliability, rather the quality of data being put into the system. Taking this point one step further, we can extrapolate a potential opportunity for IoT integration by a third party. Interviewee #6 points out that many established players already have reliable tracking infrastructure that serves their needs well enough. One of the oracle projects interviewed started as a company experimenting with on-chain supply chain tracking and pivoted to become an oracle company after recognizing a more general need for verifiable data.

Another specific use case identified over the course of the interviews is carbon certificates. Carbon certificates, or off-sets, have long had issues undermining their effectiveness. The topic of on-chain carbon certificates was discussed in interviews 2, 4, 5, and 8. In theory, tokenized carbon credits have several benefits as opposed to off-chain carbon certificates, primarily with respect to transparency and accountability. As with the supply chain case discussed previously, there is an issue regarding the quality of input data. Whereas in supply chain applications, there is a risk of parties falsifying data, the tokenization of existing off-chain carbon credits brings with it reliability issues inherent to these types of credits. Interviewee #5 points out that there are "a lot of discrepancies and scrutiny in the market right now." Hyphen.Global's approach involves using deterministic data to validate climate impacts using sources such as satellite imagery, private sensors, and data made available by the World Meteorological Organization. Interviewee #3 spoke of his involvement in building an MVP enabling a company to track and verify their process of creating biofuel from corn using on-chain tokens.

There has generally been a push for environmental projects in the Web3 space. Regenerative Finance, also called ReFi (Schletz, 2022), includes a variety of applications and platforms centered around

 doi:10.20944/preprints202405.0218.v1

carbon offsetting and sustainability. This space could see a surge enabled by access to deterministic data provided on-chain as the decentralized finance space did.

Remaining within the realm of Environmental, Social, and Governance (ESG), price data regarding individual goods and services and CPI data allow for interesting use cases. Interviewee #12 specifically mentioned the concept of currencies backed directly by individual commodities or pegged to a basket of goods. Ampleforth, discussed earlier in the paper, was brought up as one such experiment. Generally speaking. However, Interviewee #12 conceded that innovation will need to compromise with regulation. "I would love to see disruptive change. I fear the regulatory environment is going to force everything to become a lot more incremental change."

Overall, it appears as though the market is largely waiting for regulatory clarity with regard to representing real-world assets on-chain. Uncertainty is holding large institutions back, while the oracle market is currently largely DeFi-driven. Nonetheless, there are experimental projects building infrastructure likely to remain unaffected by future regulation and projects operating in grey areas that may or may not succeed in the long term, such as perpetual futures protocols. Ultimately, the true potential for oracles as a whole lies in connecting off-chain and on-chain, not just in price feeds.

5.3.3. Innovation Drivers

The practical applications described above are just some examples that have been, or are currently, being experimented with. Furthermore, interviewees appeared excited by other opportunities that are, so far, more theoretical. Yet these factors are expected by our interviewees to play a significant role in the adoption of oracles and the future hybrid smart contract space.

Interviewees 3 and 8 highlighted efficiency gains by adopting blockchain-based systems as a primary driver for adoption. Interviewee #8 points out that larger institutions face significant inefficiencies that could easily be overcome by switching to automated on-chain systems. Specific examples included the settlement times of stocks in traditional markets. Interviewee #3 agrees and adds, "You could probably save billions of dollars by automating things". Interviewee #3 expects to see minor efficiency gains at large institutions as a priority before the creation of newer smaller markets, as these minor gains in systems processing billions amount to significant revenue opportunities - though they are less exciting to the interviewee when compared to novel applications of this new technology.

Interviewee #8 points out that innovation around scalability and security, such as zero-knowledge proofs and rollups, enhances the ability to bring more data on-chain securely. A special emphasis is put on the potential significance of zero-knowledge proofs, as implemented by Chainlink's DECO product. The interviewee gives an example of use cases for retail users, such as being able to verify age without revealing one's date of birth or allowing users to prove to a DeFi protocol that their credit score off-chain is above a certain threshold in order to take out an undercollateralized loan without revealing details such as bank account number, name, address, account balance, etc. On the other hand, the interviewee sees potential for zero-knowledge proofs for large institutions as well.

When trading firms wish to purchase a significant amount of a synthetic asset, they can use zero-knowledge proofs to demonstrate their financial capacity without revealing exact amounts. For instance, if a firm plans to buy $1 million worth of an asset, they can prove they have at least $2 million in their bank account without disclosing the precise balance. Similarly, if another firm wants to borrow 1,000 Tesla stocks, a lender can confirm they possess at least 1,000 Tesla stocks available for lending without specifying the total number of Tesla stocks they own. Regulatory requirements mandate these firms to maintain the confidentiality of their holdings. This principle applies to almost every major trading institution worldwide. According to interviewee #8, "If you can solve the scalability and the privacy thing, then those are the two major barriers to getting all of finance on-chain".

In the case of hybrid smart contract adoption, interviewee #12 considers a comprehensive user experience to be one of the major enablers of innovation that has so far been missing. According to this interview, the future is about creating a seamless UX that integrates features like multi-party

 doi:10.20944/preprints202405.0218.v1

computation (MPC) to securely share passwords and mnemonic phrases, combined with AI capabilities. The integration of these components will revolutionize the user experience of hybrid smart contracts.

All in all, there appear to be two factors driving the adoption of hybrid smart contracts. (1) Factors pushing users toward adoption such as efficiency gains, ease of use, reliability, and trustlessness. (2) On the other hand, there needs to be a removal of barriers to adoption, such as improving cost efficiency through scaling, addressing privacy needs, and improving the user experience.

Overall, the interviews and their analysis reveal an exciting landscape of innovation, opportunity, and difficulties. The ability to connect the on-chain and off-chain worlds allows a wide range of use cases, from tokenizing physical assets to developing more efficient systems. Notably, despite the anticipation surrounding these technologies' promise, there are also inherent difficulties. Although DeFi dominates the market at the moment, there are applications to be found in fields such as supply chain management, sustainability, and more. Developments regarding the user experience, along with improvements in scalability and security, are crucial to the successful realization of this potential.

5.4. Challenges & Implementations

This section summarizes insights regarding decentralization, security, validity, and reliability in both a general context and the context of specific oracle protocols. Furthermore, several interviewees highlighted challenges regarding different types of data, and we will discuss the concept of progressive decentralization introduced earlier.

5.4.1. General Challenges & Design

Over the course of the interviews, identifying and appropriately handling unfavorable behavior was identified as a core challenge dealt with differently by different oracle protocols. As interviewee #8 points out, when it comes to the reliability and accuracy of oracle data feeds, "anything less than 100% is not acceptable". It is important for oracle implementations to have a nuanced incentive system to ensure proper behavior. Interviewee #7 highlights that simple outlier detection alone may not be sufficient, as being an outlier does not necessarily mean being malicious. Incentives and punishments need to account for this. According to Interviewee #7, the initial approach Chainlink intended to use for slashing, reducing the stake of a node deemed to be acting unfavorably, was simply to slash any node that was an outlier. The interviewee deemed this ineffective and overly aggressive. "I think you could actually mathematically prove you simply cannot rely on outlier detection for security".

Interviewee #14 disagrees, describing a scenario in which a number of nodes retrieve a piece of data from an API. The nodes aggregate the data and find a value the majority agrees on, with outliers being filtered out. Nodes whose data is filtered out lose their posted collateral, while the others receive their collateral together with a reward. In fact, several of the oracle protocols investigated in the context of this paper use such a simple mechanism. Incentive design may drastically depend on the factors of a specific use case. The mechanism of interviewee #14 may work well with data that is pulled from the same source at the same time, but then the only factor being ensured through staking and slashing is that none of the nodes tampered with the data, not accounting for incorrect data. There may be complications in the case of a DON in which the data is sourced from multiple different places or needs to be updated frequently. Interviewee #6 points out that a constant feed of data may lead to incorrect data spikes. "You cannot give constant data that is reliable". Sophisticated oracle implementations optimized for decentralization and speed require nuanced incentive design depending on their architecture.

Interviewee #15 eloquently summarizes oracles implementations to fall into three main architecture types. (1) The algorithmic model is the most common. This architecture is best suited for numeric data streams provided by multiple reliable data sources such as market data and weather data. Nodes may connect directly to a data source, aggregate data to reach consensus, and transmit the result on-chain. (2) The optimistic model relies on one network participant to post collateral and provide the result to a query. This approach is well suited for one-off data requests such as "Did it rain

 doi:10.20944/preprints202405.0218.v1

in Berlin at any time on March 12th?" or "Which team won the recent football match between Spain and France?". Once a response to such a request is submitted, a dispute period begins. If the claim remains undisputed, the responder receives their bond and a reward. The claim may also be disputed by another network participant also posting a bond, which they stand to lose, to open a dispute. These oracles are called "optimistic" because they assume the submitted data to be true, with the majority of responses going undisputed. (3) Single-source data feeds may be used in cases where there is only one centralized data source, such as the odometer of a car or the balance of a bank account. Single-source feeds may also be used in the initial implementation of experimental use cases. Single-source data feeds may present more points of failure compared to more decentralized architectures, such as a data source being faulty or corrupt. However, as interviewee #9 points out, the latter may not be as much of a concern when it comes to reputable and regulated institutions.

Ultimately, different use cases require different types of data, and different types of data have differing optimal data provision approaches. All approaches may face further challenges as there might not be an ideal architecture for a data feed unless it is purpose-built considering the circumstances of the data type and the use case. As interviewee #7 puts it:

> [if] you need an oracle one time there, and it doesn't need to be fast, and it doesn't need to respond very often, it just needs to respond once. That requires a very different architecture than something like price feeds that need accurate data to the second and are like constantly being triggered.

The other major challenge pointed out in interviews 1, 3, 4, 7, 8, 11, and 15, is the cost-effectiveness of a given oracle implementation. Bringing any piece of data on-chain requires at least one transaction. On all public smart contract networks, this transaction will cost a transaction fee, which depends on the current network usage. In the case of low-throughput, high-usage networks such as Ethereum, fees for a single transaction may exceed $50 in times of network congestion. Of course, oracle protocols are attempting to minimize transaction size and take other actions to optimize transaction costs. Thus far, however, gas costs, especially for feeds that require frequent updates, present a significant obstacle. Interviewees 1 and 11 both mention the importance of offering both push and pull oracles, while Interviewee 10 highlights the possibility of reducing transaction fees through operating a purpose-built EVM blockchain. Interviewee #7 mentions a new protocol, "Chronicle" built by the team behind MakerDAO's oracle implementation, which uses "clever cryptography" to significantly reduce gas costs.

As we have established, the oracle market is likely not a winner-take-all scenario. Considering the variables in architecture presented above, it appears likely that there will be highly specialized implementations for individual use cases rather than one plug-and-play solution. Incentives, cost, and architecture must fit the target use.

5.4.2. Transparency

Oracles and hybrid smart contracts are inherently designed to solve issues caused by a lack of trust. Interviewee #3 highlights that trust goes beyond the implemented technology, and it is important to consider human-related aspects such as comprehension, word of mouth, and social trust. While many consider hybrid smart contracts to be trustless, meaning there is no need for trust because everything is written in code and will be executed as written, not everyone understands what makes a smart contract-based system any more reliable than a traditional system. The topic of decentralization quickly becomes philosophical and abstract for those not in the know. Hence, education and transparency are important.

While researching details about specific oracle projects and their data feeds, it became clear that without a deep technical understanding, it is not possible to find clear answers to fundamental questions, such as who the data providers for a certain DON are or how exactly data is transferred from an oracle specific blockchain to a public smart contract network. Interviewee #7 also highlighted this

 doi:10.20944/preprints202405.0218.v1

issue. When speaking about Chainlink, the interviewee mentioned the team not giving clear answers about data sources, but it became apparent through statistical fingerprinting that most nodes in a given DON were using the same data source. Interviewee #11 agrees that it is challenging to conclude such details about Chainlink DONs and goes on to say that Chainlink can also be "very intransparent when it comes to pricing, business model and doing business - so there are a lot of NDAs."

Interviewee #12, who brought up the need for good UX in the context of oracles, mentions that the Chainlink marketplace is difficult to navigate, suggesting an area for improvement in user experience and interface. While it is possible to browse individual feeds, the details one can view are limited. Other oracle projects likely face similar challenges; however, Chainlink was most commonly named over the course of the interviews. Overall, it is important to realize that trust-minimized technology is not necessarily enough for end users to have trust in a protocol. Transparency and auxiliary services, both of which are lacking at the moment, could be a substantial point of differentiation for oracle projects.

5.4.3. Project-Specific Comments

Several interviewees voiced criticisms aimed at specific oracle protocols. This section highlights the most notable ones. Considering Chainlink's incumbent market leader position and it being considered by some as being the current industry standard, a majority of criticisms were aimed at their implementation and actions. It is important to note that a number of interviewees are working on competing protocols. While an interview with a Chainlink employee was conducted, permission for recording and dissemination could not be granted due to internal restrictions; hence, this interview was not taken into account.

Chainlink's practice of subsidizing node operations was brought up in differing contexts in various interviews. From the perspective of interviewee #3, this is not only legitimate but was in many ways necessary to help establish a market for oracle data in the first place. On the other hand, interviewee #13 highlights that these subsidies make competing with Chainlink very difficult for other oracle implementations. "They have a token worth billions of dollars, and they can afford to pay those gas fees. It makes competing with them very, very difficult because we cannot pay those gas prices." Stopping short of alluding to Chainlink abusing its vast funding to undercut competitors. Interviewee #7 also highlighted Chainlink's ability to subsidize operations as one of the reasons why there have not been any major competitors "How does anyone compete with someone that's offering their services for free?". Interviewee #6 appears to agree, stating a belief that while it made sense to subsidize the new market with the goal of capturing market share, Chainlink was not and is not currently providing the best product.

Interviewee #6 goes on to highlight that Chainlink was able to leverage speculators by offering the LINK token as a way to raise capital. A token that, as the interviewee states, was not needed for the operation of the protocol. "You don't need a token for an oracle, Chainlink doesn't need a token". According to the interviewee, this move to raise funds contributed significantly to Chainlink's current position. The funding "helped them to offer the service for free. And by offering your service for free, of course, you become the market leader." Interviewee #7, who was active as an advisor in the early days of Chainlink, offers further insight regarding the LINK token stating that the initial security system proposed in the Chainlink Whitepaper was overly simplistic and while the token ultimately funded a lot of research and development it was not initially needed. Even now, the interviewee notes, "I'm not convinced that the token is required for the game-theoretic security of their current system." In support of this statement, the interviewee highlights that Chainlink, in its current form, is closer to a permissioned network than to a decentralized one. More about this in the following subsection. One of the interviewees, who wishes not to be identifiable in the context of this statement, adds that Chainlink's powerful storytelling and marketing has led many, themselves included, to erroneously view Chainlink's oracle implementation as being the only valid solution to the oracle problem.

 doi:10.20944/preprints202405.0218.v1

Generally, the sentiment gathered during these interviews underscored Chainlink's dominant position in the oracle market and role in cultivating the initial market for oracle data. While its practices, particularly the subsidization of node operations and the sale of the LINK token, have been instrumental in solidifying its market position, these methods concurrently erected barriers to entry for competitors, creating potential concerns about monopolization. At the same time, several interviewees highlighted an ongoing market shift in favor of more specialized oracle solutions. These differing points of view hint at the nuances in the industry landscape as projects balance decentralization and nurturing this new market.

5.4.4. Protocol Decentralization

The point regarding decentralization and trade-offs it may bring came up in several interviews. Interviewee #11 appears to believe that achieving decentralization in the context of oracle protocols in general is relatively challenging. The interviewee also highlights that decentralization must come over time as increased decentralization may disrupt a project's ability to innovate quickly, and until a project is able to provide a polished revenue-generating product, decentralization should not be the top priority. As in the early days, the problems brought through centralization may initially be dealt with through transparency and openness. This idea of projects as a whole, all the way down to individual data feeds and use cases, moving from curated and centralized to open and decentralized, came up during several interviews and will be referred to as progressive decentralization going forward. While this progressive decentralization brings potential points of controversy, it does seem justifiable and even necessary in certain cases.

Previously, we identified three types of oracle implementation based on data type: algorithmic, optimistic, and single source, which differ in terms of decentralization. As interviewee #11 points out, for example, single-source oracles such as a proof of reserves feed based on an API provided by a bank showing the balance of a certain account do necessarily benefit from decentralization. Interviewee #2 points out that using a decentralized network of oracle nodes may still ensure the data provided via the API is identical to what is delivered on-chain. Optimistic oracles, which rely entirely on game-theoretical incentives for security, on the other hand, are naturally more decentralized. Interviewee #7 brings up Tellor as an example of a highly decentralized and permissionless oracle implementation, as anyone can report data in response to a certain request. Naturally, the same would hold true for other, similar implementations such as UMA. Unfortunately, optimistic oracles are not well suited for many applications, such as those requiring a consistently updating data feed. Here, an algorithmic oracle is needed. When it comes to algorithmic oracle implementations, finding the ideal level of decentralization while also making sure the project grows as intended presents a relatively subjective challenge.

Some, such as Interviewee #14, believe decentralization to be crucial for ensuring data can not be tampered with. If a hybrid smart contract ingests data that can be manipulated by an external party, it loses its security advantage. True smart contracts should be entirely uncensorable, meaning no single entity can control or change the contract's behavior. The interviewee highlights that this not only relates to data feeds but also to off-chain computation. Smart contracts can offload some computation to an external environment but must maintain the same level of censorship resistance and decentralization. Interviewee #9 offers an opposing view. According to them, not all aspects of blockchain require complete decentralization, especially where it does not provide a distinct advantage. Interviewee #4 agrees, "There's nothing magical about decentralization that makes it better than a centralized system." Implying the degree to which decentralization provides a tangible benefit depends on the individual case.

Interviewee #7 recognizes that solutions in the Oracle space range on a spectrum of decentralization, and perfect decentralization is still far from being realized. The interviewee also feels that all current solutions in the Oracle space appear "hacky", suggesting that there is room for more refined solutions in the future. The current market leader was critiqued by several of the interviewees in this

 doi:10.20944/preprints202405.0218.v1

regard. Whilst perfect decentralization may not always be possible or necessary, there appears to be an expectation for projects in the space to be striving for optimal decentralization. While several interviewees shared the view that Chainlink had some responsibility to move from curated and permissioned architecture towards transparent and decentralized systems. Interviewee #13, in particular, criticized the project, pointing out that users of Chainlink's curated data feeds were essentially trusting "a four-nine multisig" which the interviewee deemed questionable. The interviewee also mentioned Chainlink's opaque crypto-economic security model, questioning whether basic features such as slashing misbehaving stakeholders were fully implemented. When asked whether the interviewee saw Chainlink progressing towards decentralization, they replied, "they're not moving in that direction. It's, it's sort of empty promises because they've been saying it for so long."

While several of the interviewees working at various oracle implementations reasoned why their approach was optimally decentralized and provided a competitive edge, the two independent researchers represented among the interviewees highlighted Tellor and Truthcoin, respectively, as examples of well-designed oracle implementations.

Ultimately, the issue of decentralization within oracle protocols lacks a uniform resolution. While decentralization is a huge value driver in the context of hybrid smart contracts, it may be necessary and beneficial to compromise with regard to certain oracle designs. The concept of "progressive decentralization" emerges as a general trend describing how projects navigate the tension between rapid innovation and sustainability.

5.4.5. Use Case Decentralization

As mentioned previously, decentralization may not be a priority for every hybrid smart contract. As Interviewee #4 points out, the level of decentralization may differ according to the use case. A video game, for example, might not require as much decentralization as a financial application. Interviewee #3 adds that startups may only be able to utilize a centralized data source initially and then decentralize as demand grows. Interviewee #12 points out that regulation also plays an important role in deciding whether the market trends more innovative decentralized applications or permissioned systems offering incremental efficiency gains. The interviewee highlights that true disruption does not come from compromises but from innovation.

As with protocol implementations, oracle-based projects also face differing optimal decentralization levels. The priority of decentralization varies not only with the oracle design type, be it algorithmic, optimistic, or single source, but also with the particular application and developmental stage of the project. Sentiment among interviewees suggests ample room for further innovation and competition in the oracle market, with decentralization being a key element for projects to focus on.

6. Summary

Throughout this paper, the initial three guiding research questions were: (1) How do stakeholders position themselves within the competitive landscape, and what are the criticisms and advantages of their approaches? (2) How mature is the industry in its current usage, and what is its future outlook? (3) What opportunities are enabled by the industry, and what tangential businesses arise from it? These questions were elaborated upon and grouped into three main research areas: industry landscape, innovation & opportunities, and challenges & implementations. Through analyzing 15 expert interviews, this paper has highlighted the following core insights.

6.1. Conclusions

Chainlink is currently the undisputed market leader, capturing an estimated 80-90% of revenue generated in the oracle space. This can be attributed largely to the project's first-mover advantage and successful fundraising via a public token sale. Some experts raised criticisms regarding whether this token was actually necessitated by Chainlink's security model or just a way to raise funds from

 doi:10.20944/preprints202405.0218.v1

speculators. While this funding was largely used for research and development, it was also used to subsidize node operations, which made it very hard for competing projects to enter the market.

Nonetheless, interviewees overwhelmingly agreed that the oracle space is not a winner-take-all market. There are numerous competing protocols active today, many of which are specialized for specific use cases, such as price feeds. There is still plenty of space for innovation, as oracle implementations are still quite immature. Key elements of focus for projects should be decentralization, security and accuracy, latency, and efficiency in terms of gas usage and transaction costs.

Oracles are currently primarily used in decentralized finance, a market which, according to the interviewees, was made possible through oracles in the first place. Outside of DeFi, there are fewer than expected practical applications actively using oracle data. There are some experimental projects across various use cases, but market growth is relatively slow.

Interviewees showed confidence that this will change as the future role of hybrid smart contracts and oracles functions are deemed vital for future use cases, both disruptive innovative applications and iterative efficiency-enhancing applications. One example of such a future use case is tokenization, which was highlighted as a potential trillion-dollar market.

As previously mentioned, contemporary oracle implementations were largely considered to be relatively immature, with key questions regarding security models and optimal decentralization still posing uncertainties. There are potentially lucrative opportunities for oracle implementations and adjacent projects in the areas of user experience, trust, and transparency. Additionally, the need for data sourcing, analysis, and processing according to specific methodologies provides a relatively new business model. Finally, the challenge of identifying potential use cases and needed data types early presents a unique opportunity.

6.2. Practical Applications

Based on the outcomes of the expert interviews, the following general recommendations can be made. Generally, there is an expectation for oracle implementations to strive for optimal decentralization. While there are valid reasons to compromise on decentralization, it is important to address such shortcomings through transparency. Oracle protocols should implement tools to allow users to explore metrics such as DON composition, data request frequency, data sources, etc. Furthermore, the research done in connection with this paper highlights that all stakeholders, from speculators to data consumers, could stand to benefit from looking beyond the industry leader as there are many valid solutions to the oracle problem. The ideal solution is often case-dependent. Finally, there are many opportunities to contribute to oracle infrastructure, be it for new businesses entering the space or legacy businesses looking for ways to innovate. Research shows that barriers to entry in terms of cost and time are relatively low, especially for Web2 businesses seeking new ways to monetize their data, though initial earnings may be correspondingly low.

6.3. Recommendations for Future Research

This paper provides insights regarding the competitive landscape of oracle implementations, challenges faced by projects building oracle-based applications, and potential opportunities. However, several unexplored avenues which warrant further investigation were identified. First, the long-term effectiveness and sustainability of security models employed by current oracle implementations and factors regarding decentralization are vital and have not been explored in depth. Second, as the oracle market has been identified as one likely fostering specialized approaches rather than a winner-take-all market, clear identification of market segments and formal categorization of oracle solutions may be of benefit. Third, this leads to further exploration of oracle use cases outside of decentralized finance being needed. Integration of real-world assets into hybrid smart contracts, be it via direct tokenization or the creation of derivatives, appears as a priority in this scope.

 doi:10.20944/preprints202405.0218.v1

6.4. Limitations

While the research presented in this paper was conducted diligently, there are several potential limitations that warrant discussion. The use of exploratory semi-structured expert interviews inherently limits control over the conversational direction. As a result, various interviews naturally diverged, leading to a rich array of insights but also potentially missing opportunities for gathering specific information. In addition, the qualitative nature of the research raises questions about subjectivity versus objectivity and introduces the risk of bias. Many interviewees are employed in the oracle industry and thus may hold predetermined opinions or beliefs that could influence the course of the interviews, potentially affecting the validity through leading questions. Lastly, the considerable volume of data generated by the interviews opens the door for potential oversights during the data processing phase, such as minor transcription errors that could compromise the study's reliability.

Appendix A. Interviewee List

Interview	Company / Position	Interview Date	Interview Duration
#1	Team Lead - API3	31.05.2023	36:57
#2	Node Operations - Blocksize	13.06.2023	47:52
#3	Team Lead - Blockdaemon	14.06.2023	12:34
#4	Frmr. Node Operator - Blocksize	18.06.2023	25:57
#5	CEO - Hyphen Global AG	27.07.2023	22:30
#6	Independent researcher & Advisor	31.07.2023	33:04
#7	Independent researcher (Academic)	25.07.2023	37:19
#8	Node Operations - North West Nodes	09.06.2023	49:10
#9	Anonymized PoR Feed Provider	28.07.2023	28:30
#10	Executive - Pyth Data Association	27.07.2023	24:47
#11	Executive - RedStone Oracles	04.08.2023	39:11
#12	Founder - Truflation	14.06.2023	38:02
#13	Executive - Tellor	17.07.2023	32:13
#14	Founder - Witnet	21.06.2023	36:24
#15	Executive and R&D - Supra	04.08.2023	35:47

Appendix B. Interview Transcripts

Available upon request.

Appendix C. Abbreviations

NDA = Non Disclosure Agreement, UX = User Experience, CPI = Consumer Price Index

References

1. Anderson, J. 'Ike Pono—designing the political and economic systems of the Internet generation **2019**.
2. Khalil, M.; Khawaja, K.F.; Sarfraz, M. The adoption of blockchain technology in the financial sector during the era of fourth industrial revolution: a moderated mediated model. *Quality & Quantity* **2022**, *56*, 2435–2452.
3. Corradi, F.; Höfner, P. The disenchantment of Bitcoin: unveiling the myth of a digital currency. *International Review of Sociology* **2018**, *28*, 193–207.
4. Wang, X.; Ni, W.; Zha, X.; Yu, G.; Liu, R.P.; Georgalas, N.; Reeves, A. Capacity analysis of public blockchain. *Computer Communications* **2021**, *177*, 112–124.
5. Nakamoto, S. Bitcoin: A peer-to-peer electronic cash system **2008**.
6. Tapscott, A. Blockchain Revolution | Talks at Google. https://www.youtube.com/watch?v=3PdO7zVqOwc, 2016.
7. Zheng, P.; Jiang, Z.; Wu, J.; Zheng, Z. Blockchain-based decentralized application: A survey. *IEEE Open Journal of the Computer Society* **2023**.

8. Liu, J.; Liu, Z. A survey on security verification of blockchain smart contracts. *IEEE access* **2019**, *7*, 77894–77904.
9. Borge, M.; Kokoris-Kogias, E.; Jovanovic, P.; Gasser, L.; Gailly, N.; Ford, B. Proof-of-personhood: Redemocratizing permissionless cryptocurrencies. In Proceedings of the 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). IEEE, 2017, pp. 23–26.
10. Lamport, L.; Shostak, R.; Pease, M. The Byzantine generals problem. In *Concurrency: the works of leslie lamport*; 2019; pp. 203–226.
11. Jakobsson, M.; Juels, A. Proofs of work and bread pudding protocols. In Proceedings of the Secure Information Networks: Communications and Multimedia Security IFIP TC6/TC11 Joint Working Conference on Communications and Multimedia Security (CMS'99) September 20–21, 1999, Leuven, Belgium. Springer, 1999, pp. 258–272.
12. Saleh, F. Blockchain without waste: Proof-of-stake. *The Review of financial studies* **2021**, *34*, 1156–1190.
13. Solat, S.; Calvez, P.; Naït-Abdesselam, F. Permissioned vs. Permissionless Blockchain: How and Why There Is Only One Right Choice. *J. Softw.* **2021**, *16*, 95–106.
14. Athey, S.; Parashkevov, I.; Sarukkai, V.; Xia, J. Bitcoin pricing, adoption, and usage: Theory and evidence **2016**.
15. Chohan, U.W. The limits to Blockchain? Scaling vs. decentralization **2019**.
16. Liao, S. Steam No Longer Accepting Bitcoin Due to 'High Fees and Volatility'. https://www.theverge.com/2017/12/6/16743220/valve-steam-bitcoin-game-store-payment-method-crypto-volatility, 2017.
17. Uddin, M.A.; Ali, M.H.; Masih, M. Bitcoin—A hype or digital gold? Global evidence. *Australian economic papers* **2020**, *59*, 215–231.
18. Zhang, X.; Li, Y.; Sun, M. Towards a formally verified EVM in production environment. In Proceedings of the Coordination Models and Languages: 22nd IFIP WG 6.1 International Conference, COORDINATION 2020, Held as Part of the 15th International Federated Conference on Distributed Computing Techniques, DisCoTec 2020, Valletta, Malta, June 15–19, 2020, Proceedings 22. Springer, 2020, pp. 341–349.
19. Buterin, V. Devcon1: Understanding the Ethereum Blockchain Protocol. https://www.youtube.com/watch?v=gjwr-7PgpN8, 2016.
20. Khan, S.N.; Loukil, F.; Ghedira-Guegan, C.; Benkhelifa, E.; Bani-Hani, A. Blockchain smart contracts: Applications, challenges, and future trends. *Peer-to-peer Networking and Applications* **2021**, *14*, 2901–2925.
21. Presthus, W.; O'Malley, N.O. Motivations and barriers for end-user adoption of bitcoin as digital currency. *Procedia Computer Science* **2017**, *121*, 89–97.
22. Shah, K.; Lathiya, D.; Lukhi, N.; Parmar, K.; Sanghvi, H. A systematic review of decentralized finance protocols. *International Journal of Intelligent Networks* **2023**.
23. Caldarelli, G.; Rossignoli, C.; Zardini, A. Overcoming the blockchain oracle problem in the traceability of non-fungible products. *Sustainability* **2020**, *12*, 2391.
24. Beniiche, A. A study of blockchain oracles. *arXiv preprint arXiv:2004.07140* **2020**.
25. Nazarov, S. The Chainlink Network in 2023. https://blog.chain.link/the-chainlink-network-in-2023/, 2023.
26. Breidenbach, L.; Cachin, C.; Chan, B.; Coventry, A.; Ellis, S.; Juels, A.; Koushanfar, F.; Miller, A.; Magauran, B.; Moroz, D.; et al. Chainlink 2.0: Next steps in the evolution of decentralized oracle networks. *Chainlink Labs* **2021**, *1*, 1–136.
27. Palmer, M. Data is the new oil. *ANA marketing maestros* **2006**, *3*.
28. McAfee, A.; Brynjolfsson, E.; Davenport, T.H.; Patil, D.; Barton, D. Big data: the management revolution. *Harvard business review* **2012**, *90*, 60–68.
29. Marr, B. *Big data in practice: how 45 successful companies used big data analytics to deliver extraordinary results*; John Wiley & Sons, 2016.
30. Davenport, T.H.; Harris, J.G. Competing on analytics: the new science of Winning. *Harvard business review press, Language* **2007**, *15*, 24.
31. Lindecrantz, E.; Gi, M.T.P.; Zerbi, S. Personalized Experience for Customers: Driving Differentiation in Retail | McKinsey. *McKinsey & Company* **2020**.
32. Raghupathi, W.; Raghupathi, V. Big data analytics in healthcare: promise and potential. *Health information science and systems* **2014**, *2*, 1–10.
33. Baldwin, R. Netflix gambles on big data to become the HBO of streaming. *Wired. Retrieved April* **2012**, *9*, 2021.

34. Alphabet Inc.. Alphabet Inc. Form 10-K For the Fiscal Year Ended December 31, 2022. https://abc.xyz/assets/investor/static/pdf/20230203_alphabet_10K.pdf?cache=5ae4398, 2023.
35. Amazon. AMAZON.COM, INC. FORM 10-K For the Fiscal Year Ended December 31, 2022. https://d18rn0p25nwr6d.cloudfront.net/CIK-0001018724/d2fde7ee-05f7-419d-9ce8-186de4c96e25.pdf, 2023.
36. Microsoft 2022 Annual Report. https://www.microsoft.com/investor/reports/ar22/index.html, 2023.
37. Oracle Announces Fiscal 2022 Fourth Quarter and Fiscal Full Year Financial Results. https://investor.oracle.com/investor-news/news-details/2022/Oracle-Announces-Fiscal-2022-Fourth-Quarter-and-Fiscal-Full-Year-Financial-Results/default.aspx, 2023.
38. Ofulue, J.; Benyoucef, M. Data monetization: insights from a technology-enabled literature review and research agenda. *Management Review Quarterly* **2022**, pp. 1–45.
39. Melendez, S.; Pasternack, A. Here are the data brokers quietly buying and selling your personal information. *Fast Company* **2019**, *2*.
40. Giombi, K.; Viator, C.; Hoover, J.; Tzeng, J.; Sullivan, H.W.; O'Donoghue, A.C.; Southwell, B.G.; Kahwati, L.C. The impact of interactive advertising on consumer engagement, recall, and understanding: A scoping systematic review for informing regulatory science. *Plos one* **2022**, *17*, e0263339.
41. Loomis, C.J. BlackRock: The $4.3 Trillion Force. https://fortune.com/2014/07/07/blackrock-larry-fink/, 2014.
42. Haberly, D.; MacDonald-Korth, D.; Urban, M.; Wójcik, D. Asset management as a digital platform industry: A global financial network perspective. *Geoforum* **2019**, *106*, 167–181.
43. Ungarino, R. Here are 9 fascinating facts to know about BlackRock, the world's largest asset manager. https://www.businessinsider.com/what-to-know-about-blackrock-larry-fink-biden-cabinet-facts-2020-12, 2020.
44. Henderson, R.; Walker, O. BlackRock's black box: The technology hub of modern finance. *Financial Times* **2020**.
45. Bailly, A.; Blanc, C.; Francis, É.; Guillotin, T.; Jamal, F.; Wakim, B.; Roy, P. Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models. *Computer Methods and Programs in Biomedicine* **2022**, *213*, 106504.
46. Elallid, B.B.; Benamar, N.; Hafid, A.S.; Rachidi, T.; Mrani, N. A comprehensive survey on the application of deep and reinforcement learning approaches in autonomous driving. *Journal of King Saud University-Computer and Information Sciences* **2022**, *34*, 7366–7390.
47. Chui, M.; Hazan, E.; Roberts, R.; Singla, A.; Smaje, K. The economic potential of generative AI **2023**.
48. Szabo, N. The God Protocols. http://web.archive.org/web/20061230075325/http://www.theiia.org/ITAudit/index.cfm?act=itaudit.archive&fid=216, 1999.
49. Szabo, N. Unenumerated: Bit Gold. https://unenumerated.blogspot.com/2005/12/bit-gold.html, 2008.
50. Lee, D.K.C.; Guo, L.; Wang, Y. Cryptocurrency: A new investment opportunity? *Available at SSRN 2994097* **2017**.
51. Bitnodes. https://bitnodes.io/.
52. CoinGecko. Total Crypto Market Cap Chart. https://www.coingecko.com/en/global-charts.
53. Covarrubias, L.; Zadamig, J. Organization, Autonomy and Decentralization in the Information Age. *Journal of Legislation Science-Number* **2019**.
54. Ethereum.org. Ethereum Virtual Machine (EVM). https://ethereum.org/en/developers/docs/evm/, 2023.
55. Chainlink. Understanding How Data and APIs Power Next-Generation Economies. https://blog.chain.link/understanding-how-data-and-apis-power-next-generation-economies/, 2020.
56. Caldarelli, G. Real-world blockchain applications under the lens of the oracle problem. A systematic literature review. In Proceedings of the 2020 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD). IEEE, 2020, pp. 1–6.
57. Chainlink. Hybrid Smart Contracts Explained. https://chain.link/education-hub/hybrid-smart-contracts, 2023.
58. Kshetri, N. Blockchain-based smart contracts to provide crop insurance for smallholder farmers in developing countries. *IT Professional* **2021**, *23*, 58–61.
59. O'Donnell, A.R. Etherisc Launches Blockchain-Backed Parametric Flight Delay Insurance. https://iireporter.com/etherisc-launches-blockchain-backed-parametric-flight-delay-insurance/, 2022.
60. Hamledari, H.; Fischer, M. Role of blockchain-enabled smart contracts in automating construction progress payments. *Journal of legal affairs and dispute resolution in engineering and construction* **2021**, *13*, 04520038.

 doi:10.20944/preprints202405.0218.v1

61. Ghosh, R.K.; Gupta, S.; Singh, V.; Ward, P.S. Demand for crop insurance in developing countries: New evidence from India. *Journal of agricultural economics* **2021**, *72*, 293–320.
62. Rataj, E.; Kunzweiler, K.; Garthus-Niegel, S. Extreme weather events in developing countries and related injuries and mental health disorders-a systematic review. *BMC public health* **2016**, *16*, 1–12.
63. Brüntrup, M.; Heidhues, F. *Subsistence agriculture in development: Its role in process of structural change*; Grauer, 2002.
64. Ranasinghe, R.; Ruane, A.C.; Vautard, R.; Arnell, N.; Coppola, E.; Cruz, F.A.; Dessai, S.; Saiful Islam, A.; Rahimi, M.; Carrascal, D.R.; et al. Climate change information for regional impact and for risk assessment **2021**.
65. Petram, L.O.; et al. The world's first stock exchange: How the Amsterdam market for Dutch East India Company shares became a modern securities market, 1602-1700. PhD thesis, Universiteit van Amsterdam [Host], 2011.
66. Cong, L.W.; He, Z. Blockchain disruption and smart contracts. *The Review of Financial Studies* **2019**, *32*, 1754–1797.
67. Attaran, M.; Gunasekaran, A. Applications of blockchain technology in business: challenges and opportunities **2019**.
68. Semenzin, S.; Rozas, D.; Hassan, S. Blockchain-based application at a governmental level: disruption or illusion? The case of Estonia. *Policy and Society* **2022**, *41*, 386–401.
69. Young, C.R. A Lawyer's Divorce: Will Decentralized Ledgers and Smart Contracts Succeed in Cutting Out the Middleman. *Wash. UL Rev.* **2018**, *96*, 649.
70. Bush, C. Dealing with the conflicts of interest of credit rating agencies: a balanced cure for the disease. *Capital Markets Law Journal* **2022**, *17*, 334–364.
71. Merle, R. A guide to the financial crisis—10 years later. *Washington Post. https://www. washington-post. com/business/economy/a-guide-to-the-financial-crisis–10-years-later/2018/09/10/114b76ba-af10-11e8-a20b-5f4f84429666_story. html* **2018**.
72. Caldarelli, G.; Ellul, J. The blockchain oracle problem in decentralized finance—a multivocal approach. *Applied Sciences* **2021**, *11*, 7572.
73. Chainalysis. Wormhole Hack: Lessons from the Wormhole Exploit. https://www.chainalysis.com/blog/wormhole-hack-february-2022/, 2022.
74. Circle. Cross-Chain Transfer Protocol (CCTP) | Circle. https://www.circle.com/en/cross-chain-transfer-protocol.
75. Chainlink. Verifiable Randomness for Blockchain Smart Contracts. https://blog.chain.link/chainlink-vrf-on-chain-verifiable-randomness/, 2020.
76. Scott, S.V.; Zachariadis, M. *The Society for Worldwide Interbank Financial Telecommunication (SWIFT): Cooperative governance for network innovation, standards, and community*; Taylor & Francis, 2014.
77. Swift. Swift Explores Blockchain Interoperability to Remove Friction from Tokenised Asset Settlement. https://www.swift.com/news-events/news/swift-explores-blockchain-interoperability-remove-friction-tokenised-asset-settlement, 2023.
78. Sazandrishvili, G. Asset tokenization in plain English. *Journal of Corporate Accounting & Finance* **2020**, *31*, 68–73.
79. Kim, S. Fractional ownership, democratization and bubble formation-the impact of blockchain enabled asset tokenization **2020**.
80. He, S.; Manela, A.; Ross, O.; von Wachter, V. Fundamentals of perpetual futures. *arXiv preprint arXiv:2212.06888* **2022**.
81. Ng, T.S. Blockchain and beyond: smart contracts. *Bus. L. Today* **2017**, p. 1.
82. Kamath, R. Food traceability on blockchain: Walmart's pork and mango pilots with IBM. *The Journal of the British Blockchain Association* **2018**, *1*.
83. PYMNTS. Walmart Puts Pricey Blockchain Food Tracking Platform on Ice. https://www.pymnts.com/blockchain/2022/walmart-puts-pricey-blockchain-food-tracking-platform-on-ice/, 2022.
84. Caldarelli, G. Understanding the blockchain oracle problem: A call for action. *Information* **2020**, *11*, 509.
85. Meinert, E.; Alturkistani, A.; Foley, K.A.; Osama, T.; Car, J.; Majeed, A.; Van Velthoven, M.; Wells, G.; Brindley, D.; et al. Blockchain implementation in health care: Protocol for a systematic review. *JMIR research protocols* **2019**, *8*, e10994.

86. Zeiselmair, A.; Steinkopf, B.; Gallersdörfer, U.; Bogensperger, A.; Matthes, F. Analysis and Application of Verifiable Computation Techniques in Blockchain Systems for the Energy Sector. *Frontiers in Blockchain* **2021**, *4*, 725322.
87. Latifi, S.; Zhang, Y.; Cheng, L.C. Blockchain-based real estate market: One method for applying blockchain technology in commercial real estate market. In Proceedings of the 2019 IEEE international conference on blockchain (blockchain). IEEE, 2019, pp. 528–535.
88. Caldarelli, G.; Zardini, A.; Rossignoli, C. Blockchain adoption in the fashion sustainable supply chain: Pragmatically addressing barriers. *Journal of Organizational Change Management* **2021**, *34*, 507–524.
89. Chipolina, S. Oracle exploit sees $89 million liquidated on compound. *Decrypt* **2020**.
90. Qin, K.; Zhou, L.; Gamito, P.; Jovanovic, P.; Gervais, A. An empirical study of defi liquidations: Incentives, risks, and instabilities. In Proceedings of the Proceedings of the 21st ACM Internet Measurement Conference, 2021, pp. 336–350.
91. Chainlink. Three Years on Mainnet. https://blog.chain.link/three-years-on-mainnet/, 2022.
92. Chainlink. How Chainlink Price Feeds Secure the DeFi Ecosystem. https://blog.chain.link/chainlink-price-feeds-secure-defi/, 2022.
93. Berger, B.; Huber, S.; Pfeifhofer, S. OraclesLink: An architecture for secure oracle usage. In Proceedings of the 2020 Second International Conference on Blockchain Computing and Applications (BCCA). IEEE, 2020, pp. 66–72.
94. SmartContent. Band Protocol and Chainlink: A Comparative Analysis. https://smartcontentpublication.medium.com/a-comparative-analysis-of-band-protocol-and-chainlink-54b7d14823b5, 2021.
95. Band. Band Protocol | LinkedIn | Team. https://th.linkedin.com/company/band-protocol.
96. Benligiray, B.; Milic, S.; Vänttinen, H. Decentralized apis for web 3.0. *API3 Foundation Whitepaper* **2020**.
97. API3. Api3 | linkedin | Team. https://www.linkedin.com/company/api3.
98. API3. What are first-party oracles? | Documentation. https://dapi-docs.api3.org/explore/introduction/first-party.html.
99. Pyth Network: A First-Party Oracle. https://pyth.network/whitepaper.pdf, 2022.
100. Pyth Data Association. https://pyth.network/.
101. Pyth. Pyth Network | LinkedIn | Team. https://ch.linkedin.com/company/pyth-network.
102. RedStone. https://app.redstone.finance/.
103. Redstone Docs. https://docs.redstone.finance.
104. RedStone. Introducing RedStone. https://medium.com/@RedStone_Finance/introducing-redstone-1b79875df4f0, 2022.
105. RedStone. Redstone Oracles | LinkedIn | Team. https://ch.linkedin.com/company/redstone-finance.
106. Tellor. The Tellor Whitepaper | Decentralized Oracle Protocol. https://tellor.io/whitepaper/, 2022. Accessed on 2024-03-20.
107. Tellor. Tellor | LinkedIn | Team. https://www.linkedin.com/company/tellorinc/about/.
108. Core, T. Tellor Launches to Mainnet! https://tellor.io/tellor-launches-to-mainnet/, 2019. Accessed on 2024-03-20.
109. Witnet. Witnet Oracle Docs. https://docs.witnet.io/. Retrieved August 26, 2023.
110. Sánchez de Pedro, A. Witnet After Mainnet: This Is Just the Beginning! https://medium.com/witnet/witnet-after-mainnet-this-is-just-the-beginning-e4b27485c36b, 2020.
111. Dia App | Cross-Chain Oracles for Web3. https://www.diadata.org/app.
112. DIA Technical Structure. https://docs.diadata.org/introduction/dia-technical-structure.
113. Coinmonks. Dia—Open Source Oracles for Web3. https://medium.com/coinmonks/dia-open-source-oracles-for-Web3-c873ddd46a50, 2022.
114. DIA. Dia Association | LinkedIn | Team. https://ch.linkedin.com/company/diadata-org.
115. Supra Research. Dora: Distributed Oracle Agreement. https://supraoracles.com/docs/SupraOracles-DORA-Whitepaper.pdf, 2023.
116. Supra. Supra Oracles | LinkedIn | Team. https://www.linkedin.com/company/supraoracles/people/.
117. Supra Oracles. https://supraoracles.com/.
118. UMA. https://uma.xyz. Retrieved August 26, 2023.
119. UMA. Uma | LinkedIn | Team. https://www.linkedin.com/company/umaproject. Retrieved August 26, 2023.

 doi:10.20944/preprints202405.0218.v1

120. Lambur, H. Introducing UMA's Optimistic Oracle. https://medium.com/uma-project/introducing-umas-optimistic-oracle-d92ce5d1a4bc, 2021.
121. Lesaege, C.; George, W.; Ast, F. Kleros: Long Paper v2. 0.2. *Kleros* **2021**.
122. Kleros. Kleros. https://kleros.io/.
123. Kleros. Kleros | LinkedIn | Team. https://fr.linkedin.com/company/kleros.
124. Augur. The Augur White Paper: A Decentralized Oracle and Prediction Market Platform **2018**.
125. Ramirez, B. A Resurgent Online Betting Market Is Boosted by Crypto and Current Events. https://www.nbcnews.com/tech/internet/polymarket-online-bet-submersible-russia-war-rcna93122, 2023.
126. Tal, Y.; Ramirez, B.; Pohlmann, J. The Graph: A Decentralized Query Protocol for Blockchains. *Disponibile all'indirizzo: https://raw. githubusercontent. com/graphprotocol/research/master/papers/whitepaper/the-graph-whitepaper. pdf* **2018**.
127. Kaleem, M.; Shi, W. Demystifying pythia: A survey of chainlink oracles usage on ethereum. In Proceedings of the Financial Cryptography and Data Security. FC 2021 International Workshops: CoDecFin, DeFi, VOTING, and WTSC, Virtual Event, March 5, 2021, Revised Selected Papers 25. Springer, 2021, pp. 115–123.
128. Aquilina, M.; Frost, J.; Schrimpf, A. Decentralised finance (DeFi): a functional approach. *Available at SSRN 4325095* **2023**.
129. DeFiLlama. https://defillama.com/.
130. Chainlink. The Chainlink Economics 2.0 Staking Protocol and Staking v0.1 Launch Details. https://blog.chain.link/chainlink-staking-launch-details/, 2022.
131. Chainlink Data. Mainnet Data. https://data.chain.link/ethereum/mainnet.
132. Chainlink Labs Research. Deco Research Series #1: Introduction. https://blog.chain.link/deco-introduction/, 2023.
133. Papacharissiou, H. Build a Parametric Insurance Smart Contract with Chainlink. https://blog.chain.link/parametric-insurance-smart-contract/, 2020.
134. Chainlink Blog. How to Connect a Tesla to a Smart Contract via a Chainlink Node. https://blog.chain.link/create-tesla-smart-contract-rental/, 2020.
135. Stipčević, M.; Koç, Ç.K. True random number generators. In *Open Problems in Mathematics and Computational Science*; Springer, 2014; pp. 275–315.
136. Solouki, M.; Bamakan, S.M.H. An in-depth insight at digital ownership through dynamic NFTs. *Procedia Computer Science* **2022**, *214*, 875–882.
137. Hafid, A.; Hafid, A.S.; Samih, M. Scaling blockchains: A comprehensive survey. *IEEE access* **2020**, *8*, 125244–125262.
138. Chainlink. What is Off-chain Data and Off-chain Computation? https://chain.link/education-hub/off-chain-data, 2023.
139. Conti, M.; Kumar, E.S.; Lal, C.; Ruj, S. A survey on security and privacy issues of bitcoin. *IEEE communications surveys & tutorials* **2018**, *20*, 3416–3452.
140. Finck, M. Blockchains and data protection in the European Union. *Eur. Data Prot. L. Rev.* **2018**, *4*, 17.
141. Zhang, F.; Cecchetti, E.; Croman, K.; Juels, A.; Shi, E. Town crier: An authenticated data feed for smart contracts. In Proceedings of the Proceedings of the 2016 aCM sIGSAC conference on computer and communications security, 2016, pp. 270–282.
142. Juels, A. Town Crier and Chainlink: Enriching the Function of Blockchain Oracles. https://blog.chain.link/town-crier-and-chainlink/, 2019.
143. Jessel, B. Chainlink's New Acquisition from Cornell University Could Transform Blockchain for Good. https://www.forbes.com/sites/benjessel/2020/08/29/chainlinks-new-acquisition-from-cornell-university-could-transform-blockchain-for-good/, 2020.
144. Zhao, Y.; Kang, X.; Li, T.; Chu, C.K.; Wang, H. Toward trustworthy defi oracles: past, present, and future. *IEEE Access* **2022**, *10*, 60914–60928.
145. Truflation. Methodology. https://truflation.com/methodology. Retrieved August 26, 2023.
146. Kessler, S. Chainlink 'Proof of Reserve' Proves Little Beyond Data Going In, Coming Out. https://www.coindesk.com/tech/2023/07/05/chainlink-proof-of-reserve-proves-little-beyond-data-going-in-coming-out/, 2023.
147. LinkRiver. Guardians of the Web3 Economy: Chainlink Node Operators. https://blog.linkriver.io/chainlink-node/, 2023.

148. Mühlberger, R.; Bachhofner, S.; Castelló Ferrer, E.; Di Ciccio, C.; Weber, I.; Wöhrer, M.; Zdun, U. Foundational oracle patterns: Connecting blockchain to the off-chain world. In Proceedings of the Business Process Management: Blockchain and Robotic Process Automation Forum: BPM 2020 Blockchain and RPA Forum, Seville, Spain, September 13–18, 2020, Proceedings 18. Springer, 2020, pp. 35–51.
149. Ampleforth docs. https://docs.ampleforth.org/.
150. Murimi, R.M.; Wang, G.G. On elastic incentives for blockchain oracles. *Journal of Database Management (JDM)* **2021**, *32*, 1–26.
151. Ellis, S.; Juels, A.; Nazarov, S. Chainlink: A decentralized oracle network. *Retrieved March* **2017**, *11*, 1.
152. De Collibus, F.M.; Partida, A.; Piškorec, M.; Tessone, C.J. Heterogeneous preferential attachment in key ethereum-based cryptoassets. *Frontiers in Physics* **2021**, *9*, 720708.
153. Egberts, A. The oracle problem-an analysis of how blockchain oracles undermine the advantages of decentralized ledger systems. *Available at SSRN 3382343* **2017**.
154. Eberhardt, J.; Tai, S. On or off the blockchain? Insights on off-chaining computation and data. In Proceedings of the Service-Oriented and Cloud Computing: 6th IFIP WG 2.14 European Conference, ESOCC 2017, Oslo, Norway, September 27-29, 2017, Proceedings 6. Springer, 2017, pp. 3–15.
155. Xu, X.; Weber, I.; Staples, M. *Architecture for blockchain applications*; Springer, 2019.
156. Al-Breiki, H.; Rehman, M.H.U.; Salah, K.; Svetinovic, D. Trustworthy blockchain oracles: review, comparison, and open research challenges. *IEEE access* **2020**, *8*, 85675–85685.
157. Lo, S.K.; Xu, X.; Staples, M.; Yao, L. Reliability analysis for blockchain oracles. *Computers & Electrical Engineering* **2020**, *83*, 106582.
158. Nazarov, S.; Shukla, P.; Erwin, A.; Rajput, A. Bridging the governance gap: Interoperability for blockchain and legacy systems. In Proceedings of the World Economic Forum whitepaper. https://www. weforum. org/whitepapers/bridging-the-governance-gap-interoperability-for-blockchain-and-legacy-systems, 2020.
159. Caldarelli, G. Wrapping trust for interoperability: A preliminary study of wrapped tokens. *Information* **2021**, *13*, 6.
160. Šimunić, S.; Bernaca, D.; Lenac, K. Verifiable computing applications in blockchain. *IEEE access* **2021**, *9*, 156729–156745.
161. Eskandari, S.; Salehi, M.; Gu, W.C.; Clark, J. Sok: Oracles from the ground truth to market manipulation. In Proceedings of the Proceedings of the 3rd ACM Conference on Advances in Financial Technologies, 2021, pp. 127–141.
162. Ezzat, S.K.; Saleh, Y.N.; Abdel-Hamid, A.A. Blockchain oracles: State-of-the-art and research directions. *IEEE Access* **2022**, *10*, 67551–67572.
163. API3. API3 DAO Tracker - on-chain analytics: Members, staking rewards, API3 token circulating supply. https://tracker.api3.org/.
164. Band. Band Protocol - Cross-Chain Data Oracle. https://bandprotocol.com.
165. Blockchains could breathe new life into prediction markets. *The Economist* **2018**.
166. Chainlink. Linkedin. https://www.linkedin.com/company/chainlink-labs.
167. Chainlink. A Global Team of Smart Contract Experts. https://chain.link/team.
168. Chainlink. Chainlink Developer Docs. https://docs.chain.link/.
169. Chainlink. Mixicles: Smart Contract Privacy for DeFi on Public Blockchains. https://blog.chain.link/breaking-down-mixicles-and-its-potential-to-unlock-enterprise-demand-for-defi-applications-on-public-blockchains/, 2019.
170. Pyth-Network. Governance and Staking for Pyth Tokens. https://github.com/pyth-network/governance, 2022.
171. Ranasinghe, R.; Ruane, A.; Vautard, R.; Arnell, N.; Coppola, E.; Cruz, F.; Dessai, S.; Islam, A.; Rahimi, M.; Ruiz Carrascal, D.; et al., Climate Change Information for Regional Impact and for Risk Assessment. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Masson-Delmotte, V.; Zhai, P.; Pirani, A.; Connors, S.; Péan, C.; Berger, S.; Caud, N.; Chen, Y.; Goldfarb, L.; Gomis, M.; et al., Eds.; Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2021; p. 1767–1926. https://doi.org/10.1017/9781009157896.014.
172. Schletz, M. ReFi Ecosystem Litepaper. https://www.openearth.org//blog/current-state-of-refi-a-litepaper-exploring-how-to-create-interoperability-in-the-ecosystem, 2022.
173. Core, T. Tellor launches to Mainnet! | tellor. https://tellor.io/tellor-launches-to-mainnet/, 2019.

174. U.S. Bureau of Labor Statistics. Consumer Price Index. https://www.bls.gov/cpi/.
175. Witnet. Witnet Foundation | LinkedIn | Team. https://es.linkedin.com/company/witnet. Retrieved August 26, 2023.
176. Caldarelli, G. Overview of blockchain oracle research. *Future Internet* **2022**, *14*, 175.
177. Juels, A.; Breidenbach, L.; Coventry, A.; Nazarov, S.; Ellis, S.; Magauran, B. Mixicles: Simple private decentralized finance, 2019.
178. Team, M. The Maker Protocol: MakerDAO's Mulfi-Collateral Dai (MCD) System. *White paper* **2020**, pp. 1–25.
179. Mammadzada, K.; Iqbal, M.; Milani, F.; García-Bañuelos, L.; Matulevičius, R. Blockchain oracles: A framework for blockchain-based applications. In Proceedings of the Business Process Management: Blockchain and Robotic Process Automation Forum: BPM 2020 Blockchain and RPA Forum, Seville, Spain, September 13–18, 2020, Proceedings 18. Springer, 2020, pp. 19–34.
180. Marbouh, D.; Abbasi, T.; Maasmi, F.; Omar, I.A.; Debe, M.S.; Salah, K.; Jayaraman, R.; Ellahham, S. Blockchain for COVID-19: review, opportunities, and a trusted tracking system. *Arabian journal for science and engineering* **2020**, *45*, 9895–9911.
181. Pasdar, A.; Lee, Y.C.; Dong, Z. Connect api with blockchain: A survey on blockchain oracle implementation. *ACM Computing Surveys* **2023**, *55*, 1–39.
182. PwC. Blockchain: The $5 Billion Opportunity for Reinsurers **2016**.
183. Al Sadawi, A.; Hassan, M.S.; Ndiaye, M. On the integration of blockchain with IoT and the role of oracle in the combined system: The full picture. *IEEE Access* **2022**, *10*, 92532–92558.
184. De Pedro, A.S.; Levi, D.; Cuende, L.I. Witnet: A decentralized oracle network protocol. *arXiv preprint arXiv:1711.09756* **2017**.
185. White, L.J.; et al. Credit-rating agencies and the financial crisis: Less regulation of CRAs is a better response. *Journal of international banking law* **2010**, *25*, 170.
186. Xu, X.; Pautasso, C.; Zhu, L.; Gramoli, V.; Ponomarev, A.; Tran, A.B.; Chen, S. The blockchain as a software connector. In Proceedings of the 2016 13th Working IEEE/IFIP Conference on Software Architecture (WICSA). IEEE, 2016, pp. 182–191.