Spring 7-10-2024

Artificial Intelligence and Blockchain Technology

Patrick Buckley
University of Limerick, Patrick.Buckley@ul.ie

Artificial Intelligence and Blockchain Technology

Patrick Buckley[1] Patrick.Buckley@ul.ie
[1]*University of Limerick, Limerick, Ireland,*

Abstract

Recent rapid developments in Artificial Intelligence (AI) have led many observers to believe we are on the cusp of a revolution, with AI poised to have an enormous impact upon societies and economies. However, many challenges must be met before AI can safely and fairly fulfil its potential. Blockchain is a set of inter-related, interconnected technologies that allow for the development of a range of socio-technical constructs such as data markets and prediction markets which have unique attributes and capabilities such as data immutability and designable anonymity and privacy. This research explores how these attributes and capabilities of these systems could be leveraged to address some of the challenges in AI development.

Keywords: Artificial Intelligence, Blockchain Technology, Prediction Markets, Data Markets

Introduction

Artificial Intelligence (AI) is currently experiencing a massive boom in attention from both academics and practitioners. The confluence of a number of trends including increasingly powerful hardware, the increasing availability of data in digital formats and methodological breakthroughs in areas such as deep learning and reinforcement learning has led to the deployment of several eye-catching AI applications such as ChatGPT. These developments have led many observers to believe we are on the cusp of an AI revolution.

The excitement generated by recent developments notwithstanding, there are many fundamental challenges that must be addressed before AI can fulfil its touted potential. Some are technical in nature – for example, how can we manage and validate the vast amounts of data modern AI systems require? Others are social or political in nature – how can society integrate AI decision-making systems into our businesses and societies in a manner that is efficient, just and ethical? Such questions are intimidating in scope and will undoubtedly require contributions from a wide range of fields and disciplines.

Blockchain is a set of inter-related, interconnected technologies that allow for the development of a range of socio-technical constructs such as data markets and

prediction markets which have unique attributes and capabilities such as data immutability and designable anonymity and privacy. This research aims to explore how these attributes and capabilities could be leveraged to address some of the challenges in AI development.

Literature Review

In this section, a brief overview of the two fields to be synthesized in provided. First, a high level introduction to AI is provided. The particular purpose of this section is to identify the key high level challenges that are faced by academics and practitioners seeking to design, develop and deploy AI systems. The second field surveyed is blockchain technology. In this section, focus is given to elucidating the specific capabilities offered by blockchain technology. This catalogue is used to analyse how blockchain technology can offer potential solutions to some of the key challenges slowing the progress of AI.

Artificial Intelligence

The development of machines that can mimic or surpass human intelligence has been predicted since before the dawn of digital computing. Issac Asimov's description of sentient robots and the "Three Laws of Robotics" first appeared in print in 1942 (Asimov 1950). The Dartmouth Summer Research Project on Artificial Intelligence conference organized in 1956 is generally seen as the origin of academic AI research (Nilsson 2009). The field is notoriously volatile, with sentiment in the field oscillating between euphoric over-expectations (Crevier 1993) followed by "AI winters" of pessimistic retrenchment (Nilsson 2009).

Today, AI research is experiencing a period of sustained interest and optimism. New techniques such as genetic algorithms have been developed, while techniques with a longer pedigree such as neural networks have been reinvigorated by innovative approaches such as deep learning. Along with these theoretical advanced, increasingly powerful computing platforms and the immense data sets generated by the internet have allowed AI powered systems to make advances in areas such as voice assistants and self-driving cars (Badue et al. 2021, Hoy 2018) Other fields where AI has had a significant impact include finance, medical decision support systems, recommender

systems, face recognition and machine translation (Marr 2019). AI is often described by the mainstream media as the dominant technology of the future.
Making specific forecast in such a dynamic field is notoriously difficult. A commonly selected target for such forecasts is that of an AI system demonstrating human-level general intelligence (Baum, Goertzel, and Goertzel 2011). Presenting an aggregated summary of several surveys of AI expert communities, Bostrom (2016) provides the following median estimates: A 10% probability of Artificial General Intelligence (AGI) by 2022, a 50% probability by 2040 and a 90% probability by 2075.

However, despite positive forecasts, there is still a vast degree of uncertainty about the future developmental trajectory and impact of AI (Mindell and Reynolds 2022). Some experts foresee AI systems being tasked with building even more advanced AI systems leading to a “Cambrian explosion” of intelligence (Muehlhauser and Salamon 2012). Forecasts of this nature often see AI systems with a level of intelligence comparable to human beings as being a temporary milestone along the road to systems which dramatically exceed the intellectual capacity of human beings (Bostrom 2016). However, such a perspective is far from universal (Mindell and Reynolds 2022). While acknowledging progress, they suggest the path to artificial intelligence may be far more difficult than cheerleaders suppose (Penrose and Gardner 2002). Some researchers believe that intelligence is fundamentally non-algorithmic and deterministic Turing machines will never be able to replicate intelligence (Penrose and Gardner 2002). Another, more philosophical issue is whether the concept of intelligence as an attribute associated with a singular entity is fundamentally flawed (Clark 2005). Instead, both consciousness and intelligence may be properties embedded in a larger cultural feedback loop. From this perspective, consciousness and intelligence are properties embedded in society and cannot be created absent a larger social context (Dennett 2017).

Uncertainty also dominates prognostications about the impact of AI on society. Optimists forecast that the impact of Artificial Intelligence to be positive (Kurzweil 2005). Cognitively superior AI will turbocharge the development of solutions to challenges such as resource depletion and climate change. AI systems and robots will perform the physical and cognitive tasks required to produce goods and services. Freed from the necessity of labour. Individual humans will have far more choice in

how they spend their time, be that in consuming entertainment, creative endeavours, or more traditional economically focused activities.

Pessimists also proffer potential futures where the development of AI has negative impacts. For example, some researchers suggest that an inferior intelligence will be unable to control either the capabilities or motivations of a superior one (Bostrom 2016). In the same way that, for example, a dog or cat is unable to even conceive of human motivations, the inferior intelligence of humans will be utterly unable to understand, much less control, AI's that advance beyond a certain level of cognitive capability. In this situation, some fear a future where humans become an endangered or extinct species (Joy 2000). Others fear the diminution and eventual destruction of human agency by the practical and philosophical superiority of AI systems (Harari 2016).

Even in a scenario where AI do not advance beyond a cognitive horizon that renders them beyond human control, pessimists raise serious concerns about the spread of AI (Arntz, Gregory, and Zierahn 2016). On the face of it, predictions that AI systems and robots will perform the majority of all of the labour required to meet human needs seem benign. However, even such an eventuality raises considerable questions. The decline in skills such as navigation and map-reading due to satellite navigation can be seen as an example of systems that start as question-answering "oracles" before evolving into authoritative "sovereigns", which can lead to learned helplessness (Bostrom 2016). More prosaically, in a context where economic activity is managed by AI systems, social and political power will reside with those who control the AI systems (Autor and Dorn 2013). An extrapolation of current trends which suggests increasing inequality and a future where societal power is vested in a small group of elite actors, while the majority of humanity has little or no real agency is not unreasonable (Harari 2016).

The above survey demonstrates briefly the difficulty of making accurate predictions with regard to AI, and the range of the possibilities that the development of the technology invokes. However, these challenges notwithstanding, the consensus in academic and wider society today is that AI will have a significant and increasing effect on societies and economies for the foreseeable future.

Challenges in Developing AI Systems and Models

Building AI models and systems is a technically difficult task. Despite the enormous amounts of time, money and effort which is being expended by a wide variety of powerful and well financed actors, there are many outstanding challenges in developing and deploying AI. In this section, the literature is used to identify some of these challenges at a relatively abstract level. This catalogue will be subsequently used to structure an analysis of where the unique capabilities of blockchain based technologies offers potential solutions.

Data Quality and Quantity

As such, the accuracy, effectiveness and efficiency of AI systems trained using approaches such as deep learning is determined to a large degree by the data that is used to train them. Accurate, reliable data allows these systems to learn efficiently. A common feature of most methodologies being used to develop AI systems today is that they require vast amounts of high quality data in order to allow them to be trained (Sun et al. 2017; Halevy, Norvig, and Pereira 2009).

This requirement for large amounts of high-quality data about the real world presents a number of major challenges (Busch 2014). First, obtaining and curating such data can be expensive and time-consuming. Second, biased, incorrect or incomplete data can lead to biased or inaccurate models (Zhang, Lemoine, and Mitchell 2018). AI models often inherit the biases present in their training data, which present significant challenges to their use, particularly in sensitive applications in the areas of finance, hiring and criminal justice. Unvetted training data can render AI models vulnerable to attacks where adversaries specifically craft the data presented during the training phase to cause the model to misclassify (Papernot et al. 2016).

Currently, most AI models are trained on datasets that have been scraped from the open Internet. This is because the Internet represents one of the few sources of datasets large enough to train LLM’s and other AI models. However, to a large degree this data is posted by unknown or anonymous individuals who knowledge and motives in posting are unknown and unknowable. The unvetted and unverified nature of the vast majority of this data raises significant challenges.

Regulation and Compliance

As the capability, scale and influence of AI systems increases in the economy and society at large, so will the need for a regulatory and compliance regime that can oversee the design, development and deployment of these systems. Governments, NGO's and transnational regulatory bodies will certainly move to insist on standards and accountability, particularly when AI systems become deployed in sensitive areas such as transportation, health, industrial relations and justice. What these regulatory regimes will govern, who will design them and how they will be implemented are all open questions of enormous import both to the developers of AI systems and the societies they operate in.

Data and Model Distribution

One of the major sources of anxiety with regard to the development of AI is the fear that because of the expense involved in training AI models such as ChatGPT or Google Bard, only a few actors will be able to access these models. As such, society at large has an interest in ensuring that the power associated with these models is not concentrated in the hands of few actors. Ensuring at least the possibility of widespread access to AI models involves ensuring the widespread distribution of at least two categories of data.

The first category is training data. Any actor seeking to create their own AI model will need access to large volumes of data to feed into an appropriate algorithm. Amassing this volume of training data is an expensive pursuit, both in terms of time and money. In addition, as mentioned previously, the volume of data by itself is insufficient unless consumers can be assured as to the accuracy and reliability of the data.

The second category of data that could be distributed is informational representation of AI models themselves. In this case, the widespread dissemination of AI technology is enabled by sharing the data that represents the trained model. Here, the need for computationally expensive training is removed, and the model can essentially be run as is.

In both cases, if the decision is made to make the AI widely available, then a significant question is how can we ensure the integrity of the data being shared. The distribution of large, verified and verifiable data sets has been a continuing challenge of the digital era, and is only rendered more pressing by the requirements brought about by the rise of AI (Philip Chen and Zhang 2014).

Continuous Learning and Adaption

In general, AI models are trained with large data sets. These datasets can only contain information up until the point where they were created. This means that AI models can only "know" about information up to the point in time when their dataset was created. This limitation can be clearly seen in interactions with ChatGPT where asking about, for example, events which happened after 2021 will only produce "hallucinations" (Kumar et al. 2023). Actors wishing to maintain the efficiency of their AI models face a pressing need to continually correct, update and expand their training data sets. For models which capture and scrape training data from the Internet, this does not present a significant problem, since the Internet is constantly being added to. However, data captured from the Internet has other problems. Given the pseudo anonymous open nature of the public Internet, it seems unlikely that a situation where guarantees about the validity and reliability of data gathered from it can be made. If AI models are to be built using reliable data, it seems that some mechanism from capturing and validating data will be required. This will in turn require some mechanism for rewarding not only the creation of these datasets, but also specific incentives around for truthfully validating the data and penalties for problematic data.

Ethical Considerations and Oversight

One of the biggest challenges to the widespread deployment of AI systems is significant social and political concerns about their use in automated decision-making. AI systems are already being used to make critical decisions about individuals that have ethical implications, in areas as diverse as healthcare diagnoses, autonomous vehicles, public surveillance and criminal sentencing. Moreover, most of these systems are notoriously opaque. The majority of systems, particularly those trained using deep learning methods, are essentially "black boxes" where virtually no information is provided as to how an AI system reached a particular conclusion.

Of course, it is important to note that oversight of decision processes and the accountability of decision makers are perpetual challenges, and our current systems which are largely dependent on individual humans are far from perfect. However, many of the approaches that we currently deploy to address these challenges in our current paradigms, for example mandated transparency or legal liability, seem to be ill-suited to being applied to AI systems. Ensuring ethical behaviour and accountability in AI is a complex and evolving challenge.

AI Interaction

As AI systems become more powerful and ubiquitous, there will be an increasing demand for such systems to become autonomous. It is easy to imagine AI systems tasked with performing business functions that will move beyond giving recommendations to humans and beginning to act without human oversight, if for no other reason to take advantage of the speed advantages they will have over human operators.

Such actions may be AI systems requesting information from other AI systems. Or it may take the form of an AI requesting a service from a more traditional system, such as an AI tasked with inventory control making an order with a supplier. In both cases, these interactions will raise the need for an AI system to be able to exchange value with a partner in an automated manner. Of course, such interactions are already possible, with companies, for example, providing API's to access their systems, and traditional currency being used to exchange value. However, current systems have their limitations. Traditional financial payment systems have significant limitations. They impose transactions costs, which may rise exponentially in an environment where actors are interacting at the speeds associated with digital technology. They are generally poor at handling micro-transactions, which may become increasingly common when interacting systems must pay for the computing power required for AI systems to execute. Moreover, such systems generally depend on a trust relationship existing between parties prior to transactions. All these inefficiencies may serve to diminish the productivity gains that many expect to arise from AI systems into business.

Blockchain Technology and Cryptocurrencies

A blockchain is a set of data storage units usually referred to as blocks that is stored on a list in the order in which they were created (Gorkhali, Li, and Shrestha 2020). A blockchain can be distributed, which is a storage model where copies of the blockchain are stored and synchronised across multiple computing nodes. A distributed blockchain can be used to create an unalterable database. When this database stores transactions, it allows for the creation and secure transfer of digital assets. Amongst other purposes, these digital tokens can be used to exchange value between actors, serving as what are commonly referred to as cryptocurrencies. The most famous implementation of this model to date is Bitcoin, which is a digital currency that enables users to transfer currency pseudo-anonymously without the need for a central authority regulating the transactions. Bitcoin's white paper (Nakamoto 2008) has been used as the basis of many other blockchain-based technologies.

Since the original development of the suite of technologies referred to blockchain, there has been a steady stream of theoretical and practical developments. Of particular note is the development of blockchain platforms, which seek to move beyond storing data to providing a decentralised distributed computing platform. The oldest and most prominent example of this trend, Ethereum, is a multipurpose blockchain platform. A particular feature of Ethereum is that developers can write small fragments of code, called smart contracts, which can execute on a distributed virtual machine called the Ethereum Virtual Machine (EVM) (Wood 2014) Smart contracts offer a way of digitizing and automating the execution of trustless agreements between parties (Szabo 1997). Blockchain can be viewed as a set of related, interlocking and rapidly evolving technologies that provide a set of capabilities to actors that use them. In the following section, we categorise these capabilities.

Crypto-economic Primitives

Cryptocurrencies and other applications such as NFTs are built from a suite of crypto-economic primitives, including a shared, tamperproof database or ledger, digital assets and a set of protocols that dictate how actors can interact via those primitives. There are obvious parallels between these crypto-economic primitives and the components required to create more traditional markets. The exchange of tokens or

cryptocurrencies of value can serve the same purpose as fiat currencies did in traditionally constructed markets. Similarly, the blockchain, an itemised, ever-increasing list of transactions can be trivially re-purposed into a list of exchanges made by participants, providing traceability and transparency with regard to transactions. Taken as a whole these crypto-economic primitives allow for the creation of both traditional and novel markets structure. They enable these constructions with the significantly lower overheads and efficiency improvements associated with digital environments, while also delivering the additional benefits of being shared and immutable, an important consideration in establishing and building trust in a trading environment.

Decentralisation

Blockchains such as Bitcoin or Ethereum are permissionless and public. The associated blockchain can be downloaded by anyone in the world and anybody can add records to the public blockchain. However, other models of ledger construction are possible. With permissioned blockchains only nodes that have been granted permission to access the network can download the blockchain and add records. Prominent examples of such networks include HyperLedger and Ripple. Such networks may still be decentralised, in the sense that many nodes from many different organisations in many locations may participate, and no node has a veto on adding transactions to the ledger etc. In these cases, the degree of decentralisation is a design decision in the hands of the access permission granting authority.

In generally, decentralisation lends two important characteristics to blockchain based constructs. The first is fault tolerance. The distribution of data and computing across many computers means the system as a whole has fewer points of failure. This fault tolerance is a function of the degree of decentralisation across the network as a whole. A permissionless public network like Bitcoin is essentially as resilient as the Internet itself, while, for contrast, a private blockchain consisting of nodes inside a single organisation is vulnerable to any failure that affects the entire organisation.

The second major characteristic of such systems is that the data stored on the blockchain is generally considered to be immutable, in that no single party can arbitrarily change a record once it has been added. Of course, this immutability is not

absolute. Attacks such as a 51% attack, whereby a group of malicious nodes acting together can conspire to alter the blockchain are theoretically possible. From a practical perspective however, such attacks are extremely difficult, and are again a function of the degree of decentralisation of the network, in that the more nodes that store a copy of the ledger, the harder it is to mount such attacks.

Designable Anonymity and Privacy

In the public mind, cryptocurrencies and by extension blockchain technology is often associated with anonymous and therefore legally dubious financial transactions. This is a too crude representation of the situation. In reality, blockchain technology offers a palette of design choices. This can be considered along two dimensions, that of anonymity (can the participants in a market be tied to a specific “real world” identity) and privacy (can the modifications made to data stored on a blockchain be tied to a particular participant).

At one extreme, public, permissionless blockchains like Monero essentially allows completely anonymous and private participation in a blockchain based system. Both the identity of the participant and the information they add is provably untraceable. On the other hand, blockchains may also be designed in such a way that all the transactions undertaken by a particular account are publicly visible. In this case, participants have anonymity, but not privacy. There are numerous examples of blockchains operating thusly, with Bitcoin itself being the most famous.

Other configurations are also possible. A permissioned blockchain by definition requires that participants identify themselves to a gatekeeper before they can use the blockchain and participate in the network. A permissioned blockchain can be constructed in a decentralised manner, retaining the advantages of decentralisation, while at the same time insisting that participants prove their identity. In many situations, this management of participants is a legal or regulatory necessity. However, it is also possible in this situation to construct the blockchain in such a way that transactions cannot be tied back to a particular participant. This allows for the construction of markets which are not anonymous, but are private, thereby allowing participants to add information to the blockchain without fear of social or power dynamics which can often be an impediment to truthful information revelation.

Oracles

Within their own context, blockchains are used to create an immutable ledger of irreversible transactions. These guarantees allow them to be used to exchange value in the form of Bitcoins and other crypotcurrencies. However, these guarantees only extend to data that is directly recorded on the blockchain ledger. One of the major challenge for creating blockchain and decentralised applications is that they will often require information from the “real world”. For example, to implement a simple futures contracts, two participants may agree to a smart contract that will automatically pay the second participant funds from the first participant's account if a particular stock price exceeds a particular value. The stumbling block is providing the smart contract with the stock price in the real world. Both of the participants in the smart contract have an obvious vested interest in misleading the smart contract. These misincentives can affect any third party providing information to a smart contract. This is referred to as the Oracle problem and can be simply described as the problem of gathering verified, reliable information about the real world.

This challenge is being address in a number of ways using blockchain technology. A number of approaches are being investigated. The first, and simplest, is that an independent third party is appointed as arbitrar and provider of information. This approach has the virtue of simplicity, and given a suitable third party, it is a plausible, pragmatic solution to the problem. However, it does not ultimately resolve the challenge of incentive misalignment and is contrary to the animating spirit of decentralisation. Moreover, if such a system requires human judgement, scalability will inevitably become a problem.

Other approaches seek to use the principles of decentralisation and incentive alignment. Voting is one simple solution. First is simple voting. In this model, after data has been added to the blockchain, participants are asked to vote to confirm the validity of the information. Crypto-economic primitive are used to construct additional safeguards. In order to vote, participants must stake their own cryptocurrency or equivalent digital assets on the accuracy of their vote. Participants who vote with the majority receive their own stake back, plus a percentage of the combined stakes of the participants who voted for a different evaluation of the data.

A second model is based on the notion of allowing participants to challenge an Oracle. In this case, an Oracle adds data to the blockchain. As part of adding the data, the Oracle must stake its own digital assets on the veracity of the data. After a period of time has elapsed, if no dispute is raised, then the data is confirmed and the Oracle receives a percentage fee from all blockchain participants, as payment for the information they provided. In that period of time, other participants can challenge the Oracle, by staking their own assets to contest the veracity of the provided data. If the value of the assets staked against a veracity claim exceeds a limit determined by pre-determined mathematical formula, a voting process commences, and if the Oracles outcome is rejected, the Oracles entire stake is deemed forfeit and distributed amongst the dissenters. On the other hand, if the Oracles outcome is upheld, the dissenters stakes are forfeit. This approach attempts to avoid the temporal overhead associated with simple voting, while ensuring that incentivised collective oversight applies.

Research Question

Artificial intelligence and AI systems are exciting enormous interest at this time, with both national governments and the world's largest corporations spending enormous amounts of time and money on promoting research. There are significant challenges in developing AI systems. Some of these challenges are technological in nature, but many are more concerned with the potential social, economic and political impact of AI. The breakneck speed of technological advances in this space makes the necessity of designing and developing ways of addressing these challenges all the more urgent. As such, this imperative necessitates a broad effort to draw on solutions and ideas from a wide range of disciplines and perspectives.

This paper aims to explore the question, “What challenges in building socially and economically beneficial AI systems can potentially be addressed by the application of blockchain technology?” This research is exploratory in nature. Blockchain technology allows for the creation of socio-economic artefacts that have unique properties. For example, Data Markets can be created which allow a user or participant to have certainty about attributes of a data set either stored or reference in the data market, without needing a trust based relationship with other participants in

the market. The objective of this paper is identify specific artefacts that can be created using blockchain technology that may be used to address some of the challenges raised by the development of AI.

The major intellectual work in this paper is the synthesis of two distinct disciplines, namely blockchain technology and AI, with a view to enabling new theoretical solutions and perspectives to emerge (Torraco, 2005). One of the ways that this work can be conducted is an integrative literature review (Snyder, 2019). Integrative literature reviews aim to synthesise existing mature topics in order to generate novel frameworks and new theoretical models that may advance the state of the art.

Analysis

Blockchain technology can be used to build at least three types of socio-economic constructs that have specific features and capabilities that mean they could be used to address the challenges outlined. These constructs have particular features or attributes that mean they can address some of the challenges associated with the development of AI. Such constructs can be designed to meet specific requirements. For example, because all systems built using blockchain technology allows for designable privacy and anonymity, the socio-economic constructs described can be tailored to the needs of the context.

In the following subsections, we describe three types of constructs that can be built using blockchain technology and crypto-currencies. For each type of construct, we describe how it can be used to address some of the challenges that are associated with the development and deployment of AI systems. Where appropriate, we discuss the choices available to designers that would allow them to better match systems to the socio-economic requirements.

Data Markets

The first potential application of blockchain technology to address some of the challenges associated with the development of AI systems is using blockchain technology to create data markets. In their simplest form, these would be blockchains that would store either the data used to train AI models, or the actual trained models

themselves. In either case, the data stored would have the same guarantees around immutability that are normally conferred by blockchain technology. The blockchain could be designed to match the particular balance of anonymity and privacy required. One possible concern is that the size of the datasets, particularly training datasets might be too large to be distributed in a permissionless environment. This could be addressed by either only allowing access to participants who can meet the computational requirements of storing and distributing large blockchains, or the blockchain might only store the hashs of data, rather than the data itself.

From the perspective of addressing challenges around data quality in AI development here, the attribute of a data market is that rather than imagining a solitary actor responsible for determining the accuracy or inaccuracy of information, we instead imagine an eco-system where many evaluators, potentially both human and AI interact to evaluate the accuracy of information, with successful agents being rewarded with digital assets (which in turn would have the effect of increasing their impact on future evaluations), and unsuccessful agents being penalised.

The problem of evaluating the provenience of information is similar to the challenge of constructing an Oracle that can provide access to validated real world information to a smart contract. Researchers and practitioners have developed several models on how to guarantee the integrity of data used in smart contracts on a blockchain, and these models can be applied to the problem of verifying and validating data. Broadly speaking, these models can be broken into types, those which use trusted third parties to provide data, and those which use a consensus mechanism to arrive at a evaluation of the data provided. Trusted third party models have the advantage of simplicity, but essentially serve to re-situate the validation problem. As described in the section on Oracles, consensus models attempt to use the attributes of blockchain technology to create systems where participants are incentivised to search for and reveal the most accurate evaluation of an information source they can provide.

Further mechanisms could be used to improve the evaluation of information. Evaluators could be linked to their real world identities. In this case, evaluators reliability could be tied to their skills and reputation. A person in the real world who

has an advanced qualification in Maths may be seen as a more reliable evaluator of mathematical information than someone who doesn't have a qualification.

More scalably, an alternative model would see evaluators ranked based on the combination of their history of validated evaluations and the weighting they give to their evaluation. In this case, it is easy to build AI systems who are designed to evaluate the integrity and trustworthiness of data that is presented to them. Similarly, it is easy to understand how these AI systems could interact on a market. In this model, the evaluation of information is being performed by AI's, with their advantages in speed and scale. However, rather than depending on one "black box" AI to be accurate, in this framework the accuracy of evaluation is based on a diverse group of agents who have an incentive to compete.

Using a data market to store and verify data also provides a technological foundation for building a regulatory and compliance regime. By using blockchain technology to store data, you are creating a publicly available and thus publicly reviewable data sets that can be used for training or instantiating models. By providing this information in a public, verifiable and immutable form, at least one of the pre-requisites that will be required to create robust regulatory and compliance regimes can be met.

Enabling secure, verifiable data and model distribution in another challenge associated with the development of AI systems. In the case of either the training dataset or the model itself being made available, it is necessary to be able to guarantee the integrity of the shared data. Otherwise, the significant risk is that a would be adversary would be able to corrupt the data with malicious intent, allowing them to, for example, degrade the performance of an AI system, or alter the data in such a way as to allow them to select or predict the output of the AI given certain inputs.

One of the core capabilities of blockchain technology is its ability to guarantee the integrity of data. This capability has obvious applications in the context described above. Blockchain technology can be used to verify the integrity of the both training data or model. The widespread availability of training data/models should serve as a prompt to innovation. Moreover, the distribution of training data/models could also ameliorate the potential environmental impact of the development of AI. Capturing

and storing the large data sets required for training is expensive in terms of computational power. Training models is often exponentially more expensive. In both cases, a massive amount of duplicated can be avoided, assuming state and corporate actors are willing to work collaboratively.

Prediction/Decision markets

A second potential application of blockchain technology to the address the challenges of AI is the use of blockchain based prediction markets. Prediction markets are "markets that are designed and run for the primary purpose of mining and aggregating information scattered among traders and subsequently using this information in the form of market values in order to make predictions about specific future events" (Tziralis and Tatsiopoulos 2007). This definition emphasises their use of a market mechanism to aggregate the information held by a group of participants regarding future uncertain events (Buckley 2016). It also distinguishes them from other markets, such as those whose primary purpose is investment, the hedging of risk or enjoyment (Wolfers and Zitzewitz 2004).

Since their origin in the 1980's, they have been the subject of small but steady stream of academic research. Proponents suggest that they have a number of advantages over comparable information aggregation mechanisms such as polls or expert groups. First, prediction markets encourage information revelation (Hahn and Tetlock 2006b; Hall 2010). Second, they reward participants for searching for relevant information (Berg & Rietz, 2003; Hahn & Tetlock, 2006a; Sunstein, 2006). Third, they automatically communicates and aggregate information through the use of a market (Garvey and Buckley 2010) . Another fourth benefit is that the market provides an inherent weighting mechanism for the information provided. If participants are more confident of their beliefs in a particular topic, they will be willing to buy more of the relevant contracts, and vice versa (Berg and Rietz 2006; Graefe and Weinhardt 2008; Hahn and Tetlock 2006a). Fifth, markets, particularly those implemented using information technology can scale to very large groups (Hahn and Tetlock 2006c) Rather then providing point estimates like polls, prediction markets can operate in real-time over an extended period of time (Spann and Skiera 2003). Traditionally predcition markets have been implemented using traditional computing platfomrs, but the advent of blockchain technology has excited new interest in prediction markets as the

characteristics of this technology has particular resonances with prediction markets. Prediction markets can trivially converted to what are called decision markets. In this case, rather then select from a range of possible forecasts, the market is asked to select from a range of possible decisions.

Decision markets can be a solution to address the issue of oversight and ethical considerations with autonomous AI systems making decisions. From the outset, it is important to note that this is a challenge to all forms decision-making in modern society. Corporations, governments and individual experts make poor decisions every day, and often the effects of these decisions are borne by individuals. An individual who, for example, suffers an unwarranted incarceration due to a biased decision doesn't care whether the decision is made by an AI system or a human. Poor decisions are not solely the purview of AI systems. Nonetheless, we should always seek ways to constrain the ability of individual agents, be they human, corporate or AI to make poor, malicious or short-sighted decisions.

As with Data Markets, the underlying principle here is not to seek a perfect AI decision maker free from biases or imperfections, but instead to use crypto-economic primitives to create decision markets that amalgamate the decisions of many interacting AI agents. Many of the traditional limitations that affect decision markets do not apply here. AI systems can be directed to have an opinion, and so the problem of non-participants is resolved. AI systems can interact at computational speed, and so decisions can be reached practically immediately, removing the time issues that bedevil markets that require coordination and communication in human time scales. Decision markets implicitly reward or punish participants. Other possibilities present themselves. There is no obvious technological impediment to humans participating in decision-making, allowing from human input into the decision-making process. The key point here is that what is required is a diversity of AI systems. Rather then depending on the validity and good intentions of one model and one model making actor, we are depending on the wisdom of the crowd (in this case a crowd of interacting actors, many of whom may be AI's) and a market mechanism to reward the best decision makers over time.

Smart Contracts

A third major application of blockchain technology to some of the challenges of AI development is the use of smart contracts. More advanced blockchain platforms such as Ethereum offer participants the ability to interact with smart contracts. These smart contracts are essentially programming code that represent business logic. This code executes in the context of the blockchain. They have the ability to create, store and transfer digital assets stored on the blockchain. They are guaranteed to execute in accordance with their code, and provide a way of allowing participants on a blockchain to interact in a more advanced and customised way then the simple exchange of digital assets and cryptocurrencies.

As AI systems move into the operational aspects of businesses, many expect the advantages they possess to offer increased productivity. However, these productivity gains are dependent not just on the effectiveness of the AI system themselves, but also on the systems these AIs interact with. Gains in speed and efficiency offered by an AI system may be quickly swallowed up by inefficiencies in other parts of the supply chain. In this context, the limitations of the existing financial systems in terms of handling high velocity, low value transactions amongst trustless entities may be a major impediment to the gains many expect to gain from AI use.

In this context, blockchain technology may again offer a supporting technology that can ameliorate a particular challenge to AI systems. Many blockchains are explicitly designed to support the low costs, high velocity exchange of value in a trustless environment. There is no doubt that currently operational blockchains have limitations in terms of the velocity and volume of transactions they can support, but innovations such as sharding and chaining are being actively developed to address these limitations. In addition, advanced blockchains such as Ethereum offer smart contracts, which allow for the execution of business logic securely in a trustless environment. These offer the ability to build automated marketplaces that move beyond the exchange of value. Smart contracts can be used to provide financial services such as payment processing, loans and insurance. They can also be used to provide information services and the management of digital assets. These blockchain platforms offer the potential to remove many of the inefficiencies of traditional financial networks.

Limitations and Future Research

This paper presents an integrative literature review that synthesises the research in two topics to identify spaces where blockchain technology can be used to address some of the challenges associated with AI. As such, the research is exploratory in nature, and suffers from the limitation associated with that work. The operationalisation of any of the concepts derived in this paper would require a significant body of further work, both theoretical and empirical. From a theoretical perspective, many of the concepts outlined in this paper require a more detailed examination within the proposed context. Empirically, many of the properties attributed to blockchain based socio-economic constructs, e.g. the accuracy of prediction market forecasts require would require empirical validation. The research presented here aim to serve as signposts and suggestions for where research efforts might be useful focussed in the future.

Conclusions

In this paper, we have discussed how blockchain technology can be used to build constructs that can address some of the challenges that affect the development and deployment of AI systems. The technological capabilities offered by Blockchain only represent part of the solution to these challenges. For example, blockchain based information markets theoretically allow for large numbers of actors to contribute to the creation of large sets of training data in an untrustworthy environment. This would have the effect of reducing the amount of duplicate effort actors would otherwise have to go to in order to create large data sets individually. However, for this benefit to accrue would require actors to active in collaboration. A decision market allowing AI agents to interact to arrive at a consensus necessarily requires the participation of actors who accept that their AI agents may be flawed.

More generally, many of the potential benefits of using Blockchain technology described above are predicated on actors be willing to act collaboratively. In a market economy, this is far from a given. It seems very likely that many of the actors in the AI development will be willing to forgo efficiency benefits in the name of capturing a technological edge over their competitors. Such an issue is just one of the many where the development of AI models moves into the realm of political, economic and social

considerations. The question thus becomes one of political desire for egalitarian development of AI technology and political will to enforce it. Blockchain technology is important because it provides a technological foundation that can be used to build a more egalitarian version of AI development, but it will remain dependent on a political class desiring and if necessary, forcing these more egalitarian development paths.

References

Arntz, Melanie, Terry Gregory, and Ulrich Zierahn. 2016. "The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis," May. https://doi.org/10.1787/5jlz9h56dvq7-en.

Asimov, Isaac. 1950. *I, Robot*. Garden City, N.Y.: Doubleday.

Autor, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103 (5): 1553–97. https://doi.org/10.1257/aer.103.5.1553.

Badue, Claudine, Rânik Guidolini, Raphael Vivacqua Carneiro, Pedro Azevedo, Vinicius B. Cardoso, Avelino Forechi, Luan Jesus, Rodrigo Berriel, Thiago M. Paixao, and Filipe Mutz. 2021. "Self-Driving Cars: A Survey." *Expert Systems with Applications* 165: 113816.

Baum, Seth D., Ben Goertzel, and Ted G. Goertzel. 2011. "How Long until Human-Level AI? Results from an Expert Assessment." *Technological Forecasting and Social Change* 78 (1): 185–95.

Berg, Joyce E., and Thomas A. Rietz. 2003. "Prediction Markets as Decision Support Systems." *Information Systems Frontiers* 5 (1): 79–93.

———. 2006. "The Iowa Electronic Markets: Stylized Facts and Open Issues." In *Information Markets: A New Way of Making Decisions*, edited by Robert W Hahn and Paul C Tetlock, 142–69. Washington D.C: AEI-Brookings Joint Center for Regulatory Studies.

Bostrom, Nick. 2016. *Superintelligence: Paths, Dangers, Strategies*. Reprint edition. Oxford, United Kingdom ; New York, NY: Oxford University Press.

Buckley, Patrick. 2016. "Harnessing the Wisdom of Crowds: Decision Spaces for Prediction Markets." *Business Horizons* 59 (1): 85–94. https://doi.org/10.1016/j.bushor.2015.09.003.

Busch, Lawrence. 2014. "Big Data, Big Questions| A Dozen Ways to Get Lost in Translation: Inherent Challenges in Large Scale Data Sets." *International Journal of Communication* 8 (0): 18.

Clark, Andy. 2005. “Intrinsic Content, Active Memory and the Extended Mind.” *Analysis* 65 (1): 1–11.

Crevier, Daniel. 1993. *AI: The Tumultuous History of the Search for Artificial Intelligence*. Basic Books.

Dennett, Daniel C. 2017. *From Bacteria to Bach and Back: The Evolution of Minds*. 1 edition. New York: W. W. Norton & Company.

Garvey, John, and Patrick Buckley. 2010. “Implementing Control Mutuality Using Prediction Markets: A New Mechanism for Risk Communication.” *Journal of Risk Research* 13 (7): 951–60. https://doi.org/10.1080/13669877.2010.488742.

Gorkhali, Anjee, Ling Li, and Asim Shrestha. 2020. “Blockchain: A Literature Review.” *Journal of Management Analytics* 7 (3): 321–43. https://doi.org/10.1080/23270012.2020.1801529.

Graefe, Andreas, and Christof Weinhardt. 2008. “Long-Term Forecasting with Prediction Markets A Field Experiment on Applicability and Expert Confidence.” *The Journal of Prediction Markets* 2 (2): 71–91.

Hahn, Robert W, and Paul C. Tetlock. 2006a. “A New Tool for Promoting Economic Development.” In *Information Markets: A New Way of Making Decisions*, edited by Robert W. Hahn and Paul C Tetlock, 170–94. Washington D.C: AEI-Brookings Joint Center for Regulatory Studies.

Hahn, Robert W., and Paul C. Tetlock. 2006b. *Information Markets: A New Way of Making Decisions*. Washington D.C: AEI-Brookings Joint Center for Regulatory Studies.

———. 2006c. “Introduction to Information Markets.” In *Information Markets: A New Way of Making Decisions*, 1–12. Washington D.C: AEI-Brookings Joint Center for Regulatory Studies.

Halevy, Alon, Peter Norvig, and Fernando Pereira. 2009. “The Unreasonable Effectiveness of Data.” *IEEE Intelligent Systems* 24 (2): 8–12. https://doi.org/10.1109/MIS.2009.36.

Hall, Caitlin. 2010. “Prediction Markets: Issues and Applications.” *The Journal of Prediction Markets* 4 (1): 27–58.

Harari, Yuval Noah. 2016. *Homo Deus: A Brief History of Tomorrow*. Random House.

Hoy, Matthew B. 2018. “Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants.” *Medical Reference Services Quarterly* 37 (1): 81–88.

Joy, Bill. 2000. “Why the Future Doesn’t Need Us.” *Wired Magazine* 8 (4): 238–62.

Kumar, Mukesh, Utsav Anand Mani, Pranjal Tripathi, Mohd Saalim, Sneha Roy, and Sneha Roy Sr. 2023. “Artificial Hallucinations by Google Bard: Think Before You Leap.” *Cureus* 15 (8).

Kurzweil, Ray. 2005. *The Singularity Is near: When Humans Transcend Biology*. Penguin.

Marr, Bernard. 2019. *Artificial Intelligence in Practice: How 50 Successful Companies Used AI and Machine Learning to Solve Problems*. John Wiley & Sons.

Mindell, David A., and Elisabeth Reynolds. 2022. *The Work of the Future: Building Better Jobs in an Age of Intelligent Machines*. MIT Press.

Muehlhauser, Luke, and Anna Salamon. 2012. “Intelligence Explosion: Evidence and Import.” In *Singularity Hypotheses: A Scientific and Philosophical Assessment*, edited by Amnon H. Eden, James H. Moor, Johnny H. Søraker, and Eric Steinhart, 15–42. The Frontiers Collection. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-32560-1_2.

Nakamoto, Satoshi. 2008. “Bitcoin.” *A Peer-to-Peer Electronic Cash System*.

Nilsson, Nils J. 2009. *The Quest for Artificial Intelligence*. 1 edition. Cambridge ; New York: Cambridge University Press.

Papernot, N., P. Mcdaniel, S. Jha, M. Fredrikson, Z.B. Celik, and A. Swami. 2016. “The Limitations of Deep Learning in Adversarial Settings.” In , 372–87. https://doi.org/10.1109/EuroSP.2016.36.

Penrose, Roger, and Martin Gardner. 2002. *The Emperor’s New Mind: Concerning Computers, Minds, and the Laws of Physics*. 1 edition. Oxford: Oxford University Press.

Philip Chen, C.L., and C.-Y. Zhang. 2014. “Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data.” *Information Sciences* 275: 314–47. https://doi.org/10.1016/j.ins.2014.01.015.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333–339. *https://doi.org/10.1016/j.jbusres.2019.07.039*

Spann, Martin, and Bernd Skiera. 2003. “Internet-Based Virtual Stock Markets for Business Forecasting.” *Management Science* 49 (10): 1310–26.

Sun, C., A. Shrivastava, S. Singh, and A. Gupta. 2017. “Revisiting Unreasonable Effectiveness of Data in Deep Learning Era.” In , 2017-October:843–52. https://doi.org/10.1109/ICCV.2017.97.

Sunstein, Cass R. 2006. “Deliberating Groups vs. Prediction Markets (or Hayek’s Challenge to Habermas).” *Episteme: A Journal of Social Epistemology* 3 (3): 192–213. https://doi.org/10.1353/epi.2007.0007.

Szabo, Nick. 1997. “Formalizing and Securing Relationships on Public Networks.” *First Monday*.

Tziralis, Georgios, and Ilias Tatsiopoulos. 2007. “Prediction Markets: An Extended Literature Review.” *The Journal of Prediction Markets* 1 (1): 75–91.

Torraco, R. J. (2005). Writing Integrative Literature Reviews: Guidelines and Examples. *Human Resource Development Review*, *4*(3), 356–367. https://doi.org/10.1177/1534484305278283

Wolfers, Justin, and Eric Zitzewitz. 2004. “Prediction Markets.” *The Journal of Economic Perspectives* 18 (2): 107–126.

Wood, Gavin. 2014. “Ethereum: A Secure Decentralised Generalised Transaction Ledger.” *Ethereum Project Yellow Paper* 151 (2014): 1–32.

Zhang, B.H., B. Lemoine, and M. Mitchell. 2018. “Mitigating Unwanted Biases with Adversarial Learning.” In , 335–40. https://doi.org/10.1145/3278721.3278779.