Disruptive Technologies: Opportunities & Challenges

Sriram Raghavan
IBM Research
(sriramraghavan@in.ibm.com)

1. Introduction

As the pace of transformation in the digital age continues to increase, three foundational technologies are poised to become the cornerstones of disruption. *Artificial Intelligence (AI), Blockchain*, and the *Internet of Things (IoT),* represent a triumvirate of deeply disruptive technologies that are dramatically changing the way business is conducted. These technologies are reshaping industries, professions, and entire economies at a tremendous pace, forcing businesses, governments, and individuals to grapple and make sense of their societal implications. In this milieu, a flexible balanced and forward looking regulatory and policy framework is essential to ensure responsible, ethical, safe, and fair use of these technologies without curtailing innovation and value creation.

In this paper, we will focus on two of these technologies – AI and Blockchain. We will begin with a brief overview of each technology, summarizing their potential and current level of maturity & adoption. We will then illustrate how these technologies have an interesting **dichotomous relationship** with fundamental issues surrounding ***trust, privacy,*** and ***transparency.*** *While these technologies raise new questions and challenges in these important societal dimensions, they simultaneously also promise to be important ingredients for the solutions to these challenges*.

2. Artificial Intelligence (AI)

Artificial intelligence, as a field of study within the discipline of computer science, is over sixty years old, tracing its origins to the seminal Dartmouth conference in 1956. However, over the last half a decade or more, a confluence of exponential growth in the availability of data and an exponential drop in the cost of computing power has resulted in a step change in the ability to develop and deploy practical AI systems.

2.1 State of AI

Today, we are at a point where the technology underpinnings of *narrow AI* are well understood and such systems are being deployed at scale. In this context, narrow AI refers to AI tasks that are crisp, unambiguous, and well defined, often in a single domain with data available in sufficiently large volumes to train the models that power the AI system. Applications such as language translation, speech transcription, object detection, face recognition, and many others are now being widely deployed across industries from agriculture to finance, retail, online commerce, healthcare, manufacturing, government, and others.

However, the more pervasive and disruptive impact of AI will begin to appear as we evolve from narrow AI to *broad AI,* i.e., from narrow task-specific capabilities to broader intelligence powered by richer knowledge, adaptive learning, and more powerful reasoning techniques (Figure 1). Such AI systems can do more with smaller volumes of data, learn across domains, learn across tasks, build richer internal models of specific bodies of knowledge, and apply richer forms of inference and reasoning. We are currently in the early stages of this evolution from narrow to broad AI but it is well underway and will result in AI being used to optimize decisions across industries (see Figure 2).

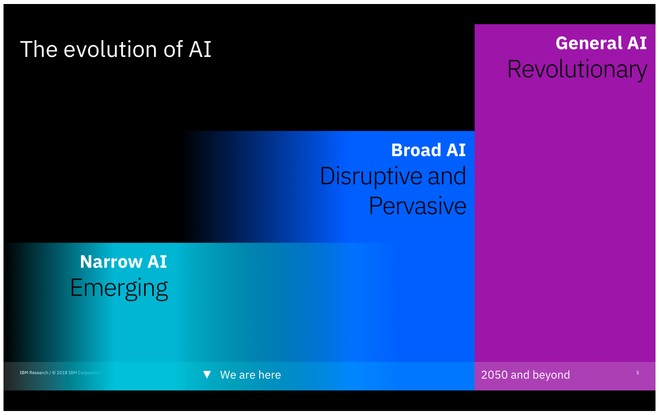
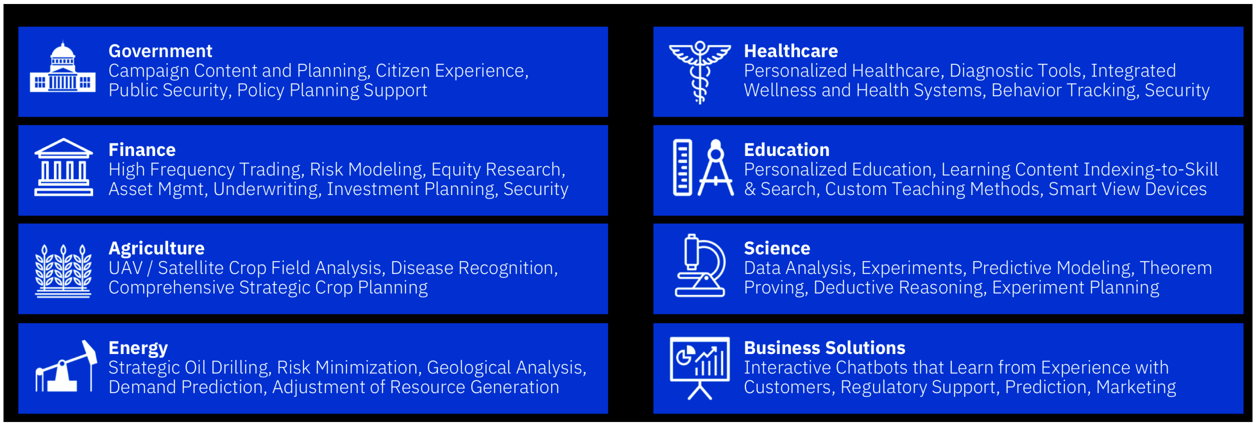


Figure 1: Evolution of AI



Government
Campaign Content and Planning, Citizen Experience, Public Security, Policy Planning Support

Healthcare
Personalized Healthcare, Diagnostic Tools, Integrated Wellness and Health Systems, Behavior Tracking, Security

Finance
High Frequency Trading, Risk Modeling, Equity Research, Asset Mgmt, Underwriting, Investment Planning, Security

Education
Personalized Education, Learning Content Indexing-to-Skill & Search, Custom Teaching Methods, Smart View Devices

Agriculture
UAV / Satellite Crop Field Analysis, Disease Recognition, Comprehensive Strategic Crop Planning

Science
Data Analysis, Experiments, Predictive Modeling, Theorem Proving, Deductive Reasoning, Experiment Planning

Energy
Strategic Oil Drilling, Risk Minimization, Geological Analysis, Demand Prediction, Adjustment of Resource Generation

Business Solutions
Interactive Chatbots that Learn from Experience with Customers, Regulatory Support, Prediction, Marketing

Figure 2: Optimizing Decisions using AI

2.2 AI Models and Data

The two fundamental building blocks of an AI system are *datasets* and *models*. Models embed the intelligence that the AI system has harnessed from processing data and are the artifacts that are embedded in business applications. For example, an application such as IBM Watson for Oncology embeds within it numerous AI models that have been trained on vast volumes of medical data related to each type of cancer.

To appreciate the issues of trust, transparency, and governance that are beginning to emerge around AI, it is important to understand the process by AI models are built from one or more data sets through a process called *training*. Figure 3 is an extremely simplified representation of the salient steps in this training process. The first phase of training is data preparation – taking raw data, often from multiple sources and in multiple forms, and using a processing pipeline to integrate, normalize, transform, clean, and curate the data into a processed data set that is ready to be utilized to build AI models. This phase of data preparation is often the most expensive, laborious, time consuming, and complex and is typically accomplished by one more data engineers. Once a prepared data set is available, a trained data scientist leverages one or more learning algorithms to build an AI model out of this data set – often experimenting with several algorithms and tuning parameters to finally result in a model that provides acceptable performance. Finally, this model is embedded into a business application where it is used to make inferences or predictions (e.g., suggest treatment regimens or drugs for a patient).

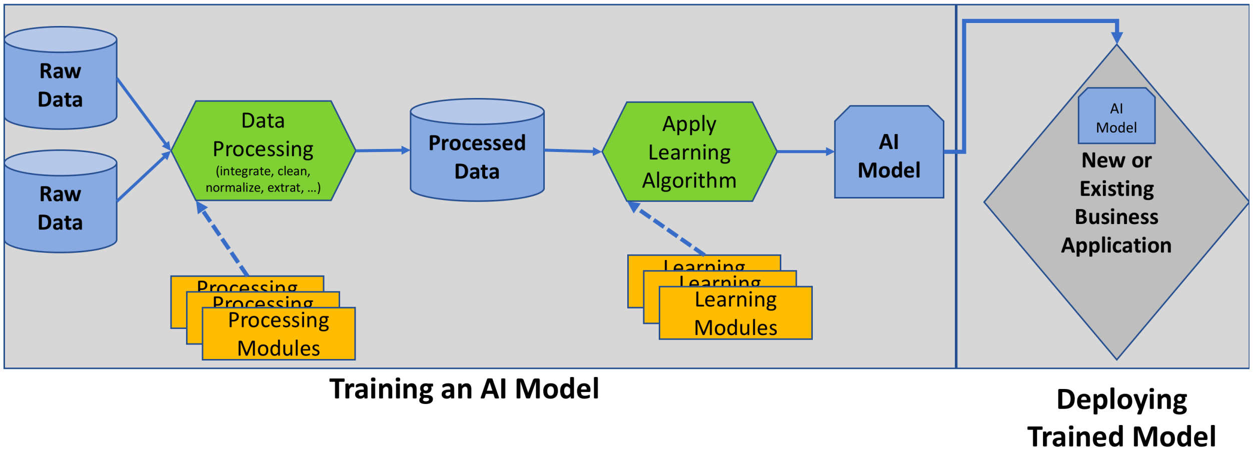


Figure 3: Central Activity in AI - Training & Deploying Models

2.3 Trust, Transparency, and Governance in AI

The process described in Figure 3 illustrates several key elements of how AI systems are built and helps draw out implications from the view point of security, trust, and transparency.

Governance of the AI Supply Chain

The models and data that make up an AI system are connected to each other through a *supply chain* analogous to physical supply chains that go from raw materials (raw data sets) to a manufactured product (AI model). Analogous to the physical world, these different data sets may belong to different entities, individual processing steps may be outsourced and performed by third parties, and individual steps of this process may take place within different national and jurisdictional boundaries. Therefore, by extension, many of the same elements that govern physical supply chains must also be asked of AI supply chains:

- *Traceability:* We must ensure that there is full trusted traceability across this supply chain so that it is possible to verify and confirm what data sets were used to train a specific AI model
- *Transparency*: There must be clarity on who trained a model, what data was used, how and where the training was accomplished, and clear historical record of the versions of all the artifacts involved in the process
- *Privacy:* As data is the fundamental raw ingredient for this supply chain, all of the issues associated with data privacy, end user control of data, and safeguarding of confidential information, also apply to the AI supply chain
- *Purpose*: Finally, policies and processes must be put in place to ensure that the purpose for which an AI model is being developed is clear & unambiguous and can be verified and audited appropriately.

Models as First Class Objects

While the digital age has introduced datasets as a fundamental asset of critical importance for a business, the AI world will add models as another artifact that should receive the same treatment. Questions of ownership, rights, trademarks, and copyrights will eventually need to be extended to cover models as well. If a company A uses the platforms, people, and infrastructure provided by company B to train a model based on a combination of its own data and data from a third-party C, how does one determine the roles responsibilities and rights of each of the entities with respect to the created model? As data sets and models become crucial elements of a company's market differentiation and competitive advantage, these questions become particularly crucial and germane.

Ethics and Bias

One of the biggest emerging questions with respect to AI systems is the question of *bias*. To understand where bias enters the AI system, it is useful to go back to Figure 3. In its very simplest form bias can emerge in three places – raw data, processed data, or model. First, bias can enter when the underlying raw data used in this AI training pipeline is biased or skewed in terms of the population that it represents (e.g., data only represents customers with a certain gender or racial background or nationality). Second, even if the original raw data sets are unbiased, bias can emerge downstream in the processing pipeline before a processed dataset is created (as mentioned earlier, the data processing pipeline is often very complex and involves many stages). Finally, even if the processed data set used to train models is unbiased, bias can enter through the training process, when statistical summaries and approximations are used to capture the model.

Public examples of the unintended consequences of bias in AI systems have already emerged. For example, there have been allegations of racism in Amazon's delivery service. In a recently published scientific paper titled "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" (Conference on Fairness, Accountability, and Transparency, February 2018), the authors evaluate multiple commercial facial recognition services and report that the capabilities are not adequately balanced for gender and skin tone. Earlier this year, as part our annual 5-in-5 technology predictions, IBM Research predicted that bias in AI systems will explode but only those coming up with solutions to monitor, control, and manage bias will survive.

The issue of recognizing, capturing, modeling, and mitigating bias is now a major theme of AI research and development activities. In the MIT-IBM Watson AI Lab, researchers are drawing upon advances in computational cognitive modeling and contractual approaches to ethics to describe the principles that humans use in decision making. IBM research scientists have recently released the world's largest annotated dataset of over 1M images for studying bias in AI systems that perform facial analysis.

The topics discussed in this section require a broad multi-disciplinary effort that combines core technical development from AI technology vendors with the right governance and legal frameworks. An example of such an initiative is the **Partnership on AI,** a premier industry wide effort to study and formulate the best practices around the development and deployment of AI technologies. The partnership was initially established in late 2016 across seven major AI technology companies – Apple, Amazon, DeepMind, Facebook, Google, IBM, and Microsoft –but has subsequently grown to over 50 partners. Many of the topics discussed in this section were the driving motivations for establishing this partnership. In addition, has IBM has released a set of **Principles for Trust and Transparency** that guide its handling of customer data and insights and the responsible development and deployment of technologies such as the IBM Watson AI system.

3. Blockchain

Blockchain technology represents the next generation of secure multi-party trusted transaction systems built on top of a shared ledger. Blockchain enables permissioned parties to come together in real time to conduct secure authenticated transactions and securely exchange data while preserving confidentiality and privacy – all without a central trusted party.

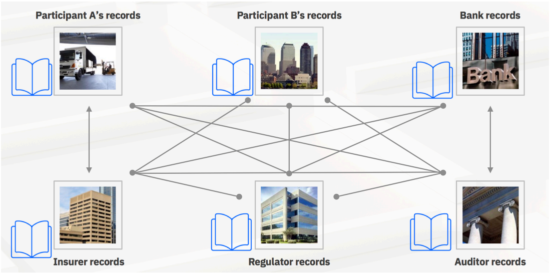


Figure 4: Inefficient business networks

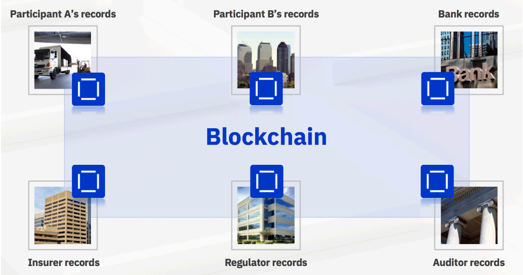


Figure 5: Business networks backed by blockchain

As a result, blockchain transforms classical inefficient expensive business networks primarily built up out of individual point-to-point connections (Figure 4) into a network backed by a shared permissioned replicated ledger holding a **single immutable version of the truth** (Figure 5)**.** This transformation results in three fundamental benefits:

- immutable version of the truth that *reduces conflict, fraud, and disputes, and the cost of reconciling inconsistent information* across individual participants
- *shared visibility of information* that drives efficiency, reduces costs, & reduces risks
- *reduction in the time to process multi-party transactions*, by removing unnecessary layers of checks, redundant operations, and duplication of work as information moves from one participant to another

3.1 State of Blockchain

While the original inspiration for blockchain technology came from anonymous cryptocurrency networks such as the bitcoin network, enterprise blockchain networks are built quite differently and operate under a different trust model. Enterprise blockchain networks are based on a *permissioned system* in which named authenticated participants come together to form a network (or join an existing permissioned network). For example, in a blockchain network for supply chain finance, the participants in the network would include the suppliers, the buyers, the banks providing finance, and potentially the logistics providers involved in moving goods from the supplier to the buyer. Similarly, a blockchain network for international trade would include shippers, customs agencies, port & terminal operators, warehouse and third party logistics providers, the exporter and importer, insurance companies, certification agencies, etc. The difference in the trust model also allows permissioned blockchain networks to operate orders of magnitude faster (1000's of transactions per second) as opposed to fully anonymous networks that typically operate at 10's of transactions per second or lower.

The ability to conduct secure authenticated transactions across a business network while recording and maintaining a single version of the truth and providing controlled visibility and sharing of data is a fundamental technological advancement with broad cross-industry applications. The most immediate impact and appetite for near term adoption & deployment is being seen in banking, financial services, and supply chain & logistics. However, numerous pilots and early-stage network formation is underway in many other industries – most notably healthcare, manufacturing, telecom, retail, government – and is expected to grow significantly over the next several years.

3.2 Governance in Blockchain Networks

As a technology that structurally and operationally cuts across many entities, it is no surprise that governance of blockchain networks is an important and complex problem. Governance cuts across many dimensions, technological and otherwise, of establishing and operating a blockchain network. It begins with the legal agreements capturing the relationships, roles, and obligations of the different participants in the network but expands to include policies for new member admission, policies around security and confidentiality of data, policies around the business rules and endorsements that control the transactions recorded on the blockchain, along with the complexities of handling regulatory compliance, risk management, and audit requirements across multiple parties.

Currently, these problems are being addressed individually in each blockchain network. However, as adoption increases, we anticipate that there will be a level of maturity and standardization that will emerge in both the legal frameworks and the associated software governance tools used to manage and operate blockchain networks.

4. The Dichotomous Relationship

The previous two sections of this paper described new issues and challenges in trust and governance posed by the emergence of AI and blockchain. Interestingly, these same technologies also provide some of the key ingredients to address these issues.

For example, we saw that one of the fundamental capabilities of blockchain networks was to introduce trust and visibility into complex multi-party transactions business networks. This same capability can be harnessed to infuse trust into the "AI supply chain" that we introduced in Section 2.2. Blockchain technology can enable improved traceability of sensitive data, as the shared immutable ledger technology brings a compelling mix of transparency and security. The decentralization of transaction processing and distribution of trust that is inherent in blockchain, the ability to store and handle encrypted data, and the ability to track and immutably record all operations, can be exploited to build trusted AI platforms that provide provenance and lineage of AI models and the entire training process. As techniques for bias checking and mitigation are invented and made part of AI platforms, blockchain can be used to record and ensure that these checks are performed and certified before AI models are deployed into production use.

Another example: while we talked about the issue of bias in AI models, AI also promises to serve as a tool to study, measure, evaluate, and address the bias inherent in society and human decision making. For example, in a recent scientific paper at the *2018 Conference on Fairness, Accountability, and Transparency,* scientists demonstrated how AI capabilities in deep natural language processing and advanced image analytics can be used to quantify and highlight gender bias in movies. Similarly, IBM's Project Debater is an early example of the power of AI to automatically extract and synthesize unbiased arguments for and against complex policy questions and potentially shine a light against society's own biases.

Finally, related to trust and governance is the important issue of data privacy, especially important as companies today collect and maintain troves of information about the personal data of their users. Recent unauthorized disclosures of sensitive personal information have re-opened fundamental questions about data privacy. These disclosures also highlight the importance of marrying policy measures with the deployment of new technologies that are designed from the ground up to safeguard privacy. Recent technical advances in areas like pervasive encryption, homomorphic encryption, secure multi-party computation, and zero knowledge protocols, while motivated by the market opportunity around blockchain, also provide the building blocks for developing secure privacy preserving systems.

5. Conclusion

We are at an inflection point in the digital age when disruptive technologies like AI and blockchain are poised to dramatically transform how individuals, society, corporations and nations function. As we embrace the immense possibilities for these technologies to transform our lives and address the most pressing challenges of our time, it is critical to establish the right frameworks, principles, and policies that ensure that these technologies are used responsibly and equitably. Clearly, this requires a strong multi-disciplinary approach and open collaboration between leaders in technology, law, policy, and government across the academic, scientific, and technical communities.