mathematics

MDPI

Review

A Survey on Big Data Technologies and Their Applications to the Metaverse: Past, Current and Future

Haolan Zhang [1,*], Sanghyuk Lee [2], Yifan Lu [1], Xin Yu [1] and Huanda Lu [1]

[1] Ningbo Institute of Technology, Zhejiang University, Ningbo 315100, China
[2] Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou 215000, China
* Correspondence: haolan.zhang@nit.zju.edu.cn

Abstract: The development of big data technologies, which have been applied extensively in various areas, has become one of the key factors affecting modern society, especially in the virtual reality environment. This paper provides a comprehensive survey of the recent developments in big data technologies, and their applications to virtual reality worlds, such as the Metaverse, virtual humans, and digital twins. The purpose of this survey was to explore several cutting-edge big data and virtual human modelling technologies, and to raise the issue of future trends in big data technologies and the Metaverse. This survey investigated the applications of big data technologies in several key areas—including e-health, transportation, and business and finance—and the main technologies adopted in the fast-growing virtual world sector, i.e., the Metaverse.

Keywords: big data; metaverse; digital human; big data technologies; virtual worlds; VR

MSC: 68T09; 62D05; 68P20; 68P27

Citation: Zhang, H.; Lee, S.; Lu, Y.; Yu, X.; Lu, H. A Survey on Big Data Technologies and Their Applications to the Metaverse: Past, Current and Future. *Mathematics* **2023**, *11*, 96. https://doi.org/10.3390/math11010096

Academic Editor: Christophe Guyeux

Received: 1 December 2022
Revised: 12 December 2022
Accepted: 20 December 2022
Published: 26 December 2022

1. Introduction

In modern society, digital applications have been extensively deployed in numerous areas. These applications can generate enormously large data, which provide abundant resources for data analytics, prediction, and decision making. The rapid growth of heterogeneous datasets demands new big data technologies for more efficient data processing. Previous research studies have provided various reviews and surveys on big data technologies. This paper reviews recent developments in big data technologies, and their applications to a virtual reality world, e.g., the Metaverse.

In the big data environment, large amounts of datasets are created, and spread rapidly, which incurs extra computational overheads [1]. Conventional data processing technologies face the challenges of data explosion, increasing data variety, limited real-time data processing efficiency and growing demands for more accurate analytical methods. The growth of data complexity and volume is hindering the application of big data analytics in the real world. In particular, e-business companies have relied excessively on online data collection and analytics: consequently, these e-business organizations seek to promote business and increase profits; nevertheless, of the e-business companies that have applied big data technologies to their business processes, only 37 percent have proven to be successful [2]. Information overload in the big data environment has increased the complexity and difficulty of the e-business decision-making process [1–3]. There is a crucial need to understand the development of big data technologies, in order to deploy more efficient new technologies to various applications.

In recent years, virtual reality applications have attracted much attention. The Metaverse is one of the most up-to-date virtual reality concepts: it is a shared three-dimensional virtual platform, which creates a mirror image of the real world based on digital twin and related technologies [4]. The use of big data technologies will unavoidably become an important issue in various virtual reality worlds, including the Metaverse. One of the main research objectives of this study was to investigate the cutting-edge big data technologies used in the Metaverse, in order to assist the development of virtual reality applications. This paper firstly explores the respective applications of big data and the Metaverse, and then connects them, in Section 4.

This paper is organized as follows: the next section provides a comprehensive survey of big data applications in different areas; Section 2 investigates the current trends in big data technologies; Section 3 reviews the development of virtual reality platforms, e.g., the Metaverse and its correlation with the big data technologies; Section 4 discusses the future application of big data technologies and the Metaverse, and their roles, and it explains the literature review methods. The last section concludes the research findings.

2. Applications of Big Data Technologies in Different Areas

'Big data' is defined as a data source that has the characteristics of large volume, high velocity, a wide range of variety, and veracity [5]. This definition is known as 'the 4Vs of big data'. In many cases, big data technology implies analytics, storage, and rapid processing in a dynamic environment. In [5], a review on medical big data applications was conducted systematically. The review summarized various applications that had been developed, based on big data technologies. This section provides a systematic review of the applications of big data technologies in several key areas.

2.1. Big Data Technologies in the e-Health Sector

The four characteristics of big data, i.e., the 4Vs, have enormous impact on medical data processing systems. Various e-health systems have been developed to tackle the problems occurring in the complex and dynamic big data environment.

Statistics indicate that the current worldwide data volume is set to more than double each year. The global healthcare data storage market is predicted to grow from USD 3.08 billion in 2020 to USD 6.12 billion by 2027 [6]. Medical data systems are facing challenges presented by the rapid growth of data volume due to the increasing use of magnetic resonance imaging (MRI), computed tomography (CT) scans, and rising patient numbers. The development of an efficient large data processing system has become a critical requirement for e-health systems.

In recent years, big-data-driven platforms for personalized healthcare have been developed, to reduce readmission rates and accelerate real-time response [7]. The extensive applications of the Clinical Data Warehouse (CDW) database have incorporated online analytical processing and sophisticated network analysis, to discover new clinical findings. In [8], a big-data-driven system was developed, to reveal and analyze hospital information system user behaviors, and potential features based on behavior analysis.

The system incorporated the extract, transform and load (ETL) module, which can process data and store it in the big data warehouse, in parallel. Like conventional large data systems, clinical systems have incorporated structure data, semi-structure data and unstructured data into their data warehouses. Most medical health systems now have the capacity to store metadata, protected health information and heterogeneous e-health data. Existing medical health systems offer cloud-based and hybrid storage solutions. Figure 1 shows the architecture of a typical health data warehouse: the medical data have structured and unstructured sources, which include patients' medical records, health monitoring signals, genome data, MRI and CT images, medical analytical reports, and laboratory testing results. It is crucial to efficiently integrate the various medical data sources in the dynamic big data environment. In [9], a patient-centric healthcare application, Health-CPS, was developed for the unification of various medical data sources. The Health-CPS system deploys data-oriented service and data collection layers with a united standard, which can efficiently integrate various medical data sources: this system demonstrates that the implementation of cloud and big data technologies in healthcare systems can enhance their system performance [5,9,10].

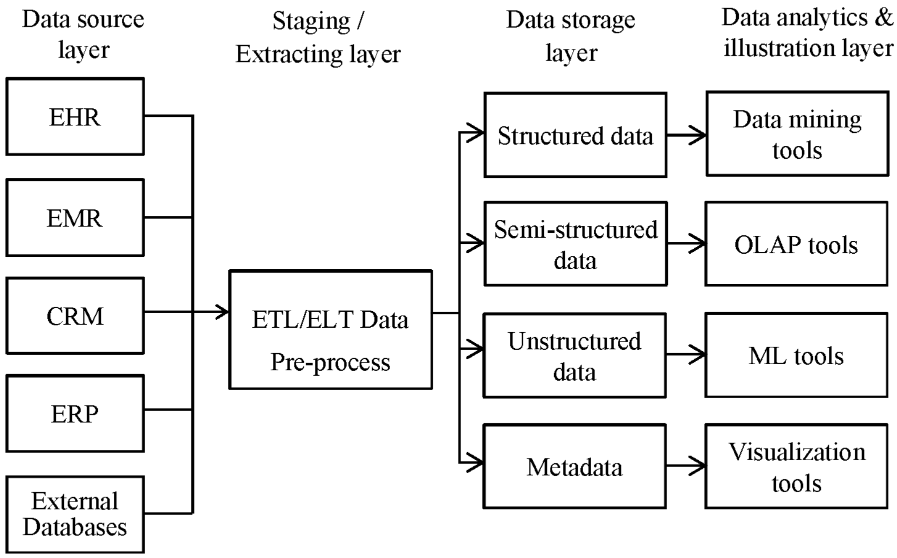

Figure 1. The architecture of a typical e-health data warehouse, modified based on [6].

From the velocity perspective, the extensive applications of wearable and sensory health monitoring devices demand a swift response, to fulfill the requirements of real-time medical data processing. Under the circumstances, the velocity of medical big data is vital to many life-saving health systems. In [11], a wearable medical emergency response system was implemented, based on a large number of medical sensors. In this system, patients' clinical data were obtained and analyzed, based on numerous wearable sensors for real-time clinical monitoring, diagnosis and treatment therapies [12].

The rapid development of wearable and sensor technologies has accelerated the applications of real-time multi-sensor wearable healthcare devices in the healthcare sector. In particular, wearable medical devices have simplified the processes of health monitoring, as they can efficiently monitor, analyze and diagnose patients' daily health conditions. In [13], a flow acceleration measurement model was developed, to assess users' posture and movement during long-term measurements in daily life: this model has been applied extensively in rehabilitation, psychophysiology and cardiology areas [5,13]. Multi-sensor medical models have focused, in recent decades, on monitoring chronic diseases. The sensor network for assessment of patients (SNAP) was developed for real-time response to patients with chronic diseases [14]. Figure 2 shows the framework of the medical SNAP model. The velocity issue is crucial in these multi-sensor wearable medical devices; therefore, more sophisticated real-time medical systems and devices have been developed, to fulfill the velocity requirement in the big data environment [15–19].

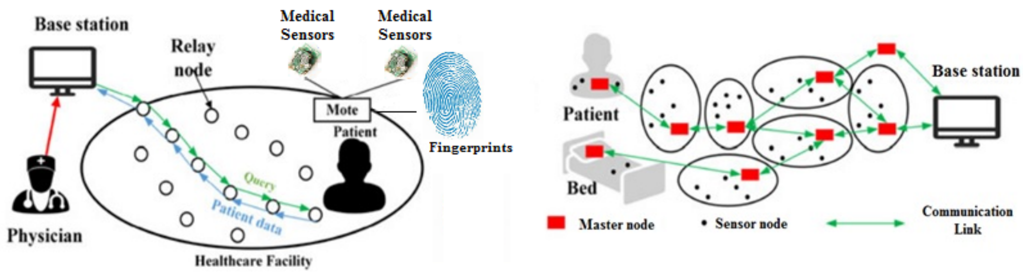

Figure 2. The framework for the medical SNAP model, modified based on [14,20].

Veracity is one of the most important characteristics, because in many cases medical data sources can be incomplete, biased and inaccurate; therefore, it is crucial to ensure data quality, in order to minimize bias, duplication, abnormalities, inconsistencies and volatility during big data analytical processes [5,21]. Research studies have shown that one of the key factors in a successful healthcare system is the implementation of efficient algorithms and data analysis technologies, to handle large amounts of heterogeneous data, so as to produce clinical results with sufficient veracity [22–24].

Previous research work has paid much attention to improving the accuracy, reliability and efficiency of medical data processing systems. Several methods have been adopted for improving the veracity of medical data sources, including data cleaning, data normalization and data fusion [25]. A new data cleaning approach has been applied for replacement of missing text, and to improve the number of relevant cases retrieved by search queries in clinical systems [26]. Some conventional methods have been used for medical data cleaning, which include [26]:

(1) Replacing missing categories, and standardizing contents in clinical reports;
(2) Abbreviations substitution, through medical dictionaries and ontologies;
(3) Filtering and eliminating data noise, errors and inconsistency, by using Natural Language Processing (NLP) methods.

In brief, veracity is undoubtedly important in medical and healthcare systems, and the development of big data technologies is improving veracity in healthcare systems.

2.2. Big Data Technologies in the Transportation Sector

Fast-growing transport systems in modern society are generating a large amount of data, including transportation trajectory data, GPS data, traffic management data, and transportation network data. Due to the increasing demands for efficient transportation systems in the big data era, various intelligent transportation systems (ITS) have been developed, to accommodate the requirements of processing and analyzing the growing amount of transport data [27–30]. Like the medical sector, transportation systems are facing the challenges of volume, velocity, variety and veracity in the dynamic big data environment. ITS applications are focusing on tackling big data problems by adopting advanced big data technologies. Research studies have shown that big data analytics can improve ITS data processing capacity, operation efficiency, and safety levels [29]. Figure 3 shows the architecture of a big data analytics framework in ITS.

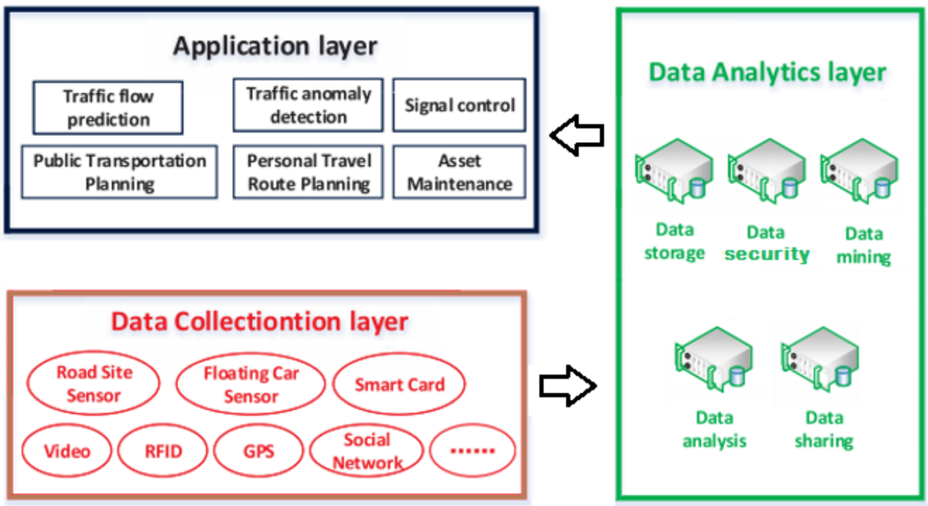

Figure 3. Big data analytics framework in ITS, modified based on [29].

The major big data analytics technologies are applied in ITS applications as follows:

- Supervised learning methods: the major data analytic and machine learning methods used in ITS include regression, decision tree, Artificial Neural Network (ANN) and Support Vector Machine (SVM) [29] (Figure 4). Linear regression is one of the most efficient methods for classification, and it has been applied extensively in ITS, for traffic route analysis and traffic flow prediction [31–33]. The decision tree method has been applied to ITS applications, such as traffic accident detection, traffic congestion prediction and accident severity prediction [34–36]. In [29], the SVM classifier with the kernel function K (x, x′) could derive the support vector α_i:

$$\max_{a_i} -\frac{1}{2}\sum_{i=1}^{l} y_i y_j \alpha_i \alpha_j K(x_i,\ x_j) + \sum_{i=1}^{l} a_i$$

$$s.t.\ \sum_{i=1}^{l} y_i a_i = 0 \tag{1}$$

where the decision function g(x) to compute the label for the sample x was:

$$g(x) = sgn\left(\sum_{i=1}^{l} y_i a_i * K(x_i, x) + b\right) \tag{2}$$

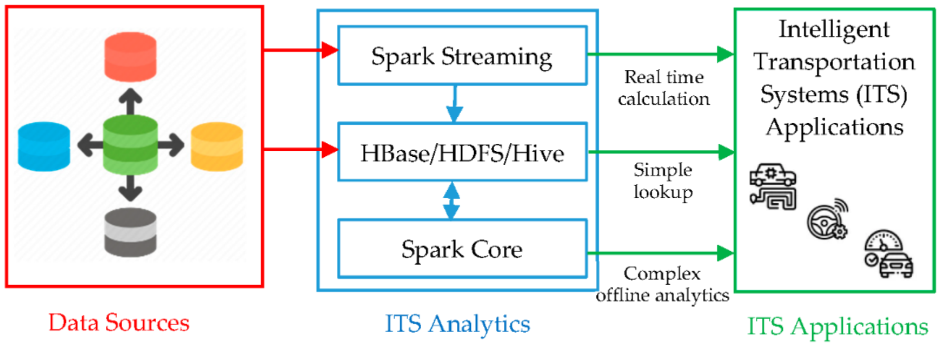

Figure 4. The use of big data platforms in ITS applications, modified based on [29].

If x is an incident sample, then $g(x)$ = 1; otherwise, $g(x)$ = −1 [29].

- Unsupervised learning and ontology-based methods: the conventional unsupervised learning method adopted in ITS is K-means, which has been applied to travel time prediction, travel path planning, etc., [29,37,38]. Ontology-based methods deploy data semantics that can efficiently associate data semantic relations, which are extensively applied in the ITS field for semantic traffic data processing [39–41].
- Deep learning and reinforced learning methods: the application of reinforced learning in ITS is to reduce the computational overhead through exploring and learning the optimal policy, based on ITS data [42]. Reinforcement learning is feasible in traffic signal control in ITS, as it incorporates supervised and unsupervised methods [43,44]. The Q-learning in reinforced learning modeling is the value iteration update, which is listed as follows:

$$Q(s_1, a_t) = Q(s_1, a_t) + \alpha\left(r_{t+1} + \gamma \max_A Q(S_{t+1},\ a_t) - Q(S_t,\ a_t)\right) \tag{3}$$

The abovementioned data analytical and machine learning methods have been used widely in the ITS area, with the support of big data platforms, such as Hadoop, Hbase, Spark, etc., [29,45].

Recently, several ITS domain-oriented big data platforms have been developed, to accommodate the rising demands for more effective data processing in ITS. A big data platform was developed in [46], with multiple engines to support heterogeneous traffic data analytics. Several other data processing platforms and frameworks have been applied to ITS, such as Godzilla [47], K-Feed [48], Sipresk [49] and the ITS big data simulation platforms [50].

In brief, big data technologies can benefit ITS applications, by enabling the highly efficient data processing and analytical capacities that contribute to the solutions of modern ITS applications, including traffic management, transportation infrastructure, transportation logistics and smart connectivity [29,51].

2.3. Big Data Technologies in the Business and Financial Sectors

Our society is entering the digital era, because of the rapid development of digital technologies. Online activities have played an important role in in our daily lives, which continuously generate large amounts of data. These data sources are generated in various sectors, including healthcare, transportation, manufacturing, finance, business, and social activities. The use of big data technologies has accelerated digital transformation processes in all sectors, enabling adaptation to market changes and the rise of the digital economy. Most modern businesses are unable to survive without the support of digital platforms [52].

The extensive application of data analytics in the commerce and finance sectors has improved efficiency in financial decision-making processes. Financial and commercial applications deploy big data technologies to perform various financial and commercial analyses, which eventually generate more accurate and comprehensive decision-making models [53–55]; however, digital applications have increased the financial risk of cyber-attack and online fraud, as personal data are rapidly becoming the new currency in the digital economy [56,57]. Technological innovation is one of the most important driving forces in financial markets, for pursuing high profits—especially digital technologies, which are applied extensively in the financial sector, and have great impact on various financial applications [54,58–61].

Modern financial systems collect accessible transparent data, to improve financial data analysis and risk control, particularly with regard to the applications of big data technologies. Big data technologies can improve the performance of both enterprise and individual-level financial systems, as they are based on large data analytics. The application of cloud computing technology in financial organizations improves data security, and reduces the costs of big data management and analysis, which facilitates the deployment of big data technologies in the business and financial sectors: for instance, Amazon Web Services (AWS) provides a scalable, cost-effective cloud platform for various businesses, including financial and commercial analytics, around the world. Figure 5 shows the big data analytics in financial services.

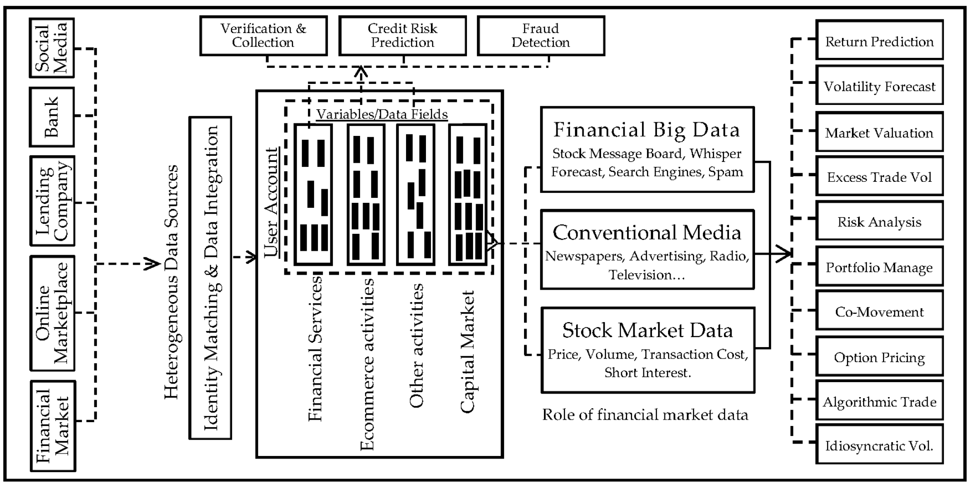

Figure 5. The framework of big data analytics in financial services, modified based on [54,58,62].

Big data technologies play a vital role in numerous online commercial activities in almost all sectors: for instance, Netflix applies big data technologies to analyzing customers' viewing behaviors, so as to provide customers with more accurate recommendations; mobile and car-based GPS applications feature location-based services such as Google Maps, which heavily rely on big data technologies for rapid customer service response [56,63,64]. Business companies utilize cloud computing to facilitate applications of big data technologies, in order to improve customer satisfaction [65]. In recent years, conventional Business-to-Consumer (B2C) companies and organizations have extended their operations to the Business-to-Business (B2B) marketing sector [63]. Consequently, more sophisticated big data technologies have been deployed in large commercial data analytics.

Fast growth, in online businesses, financial and social activities, leads to accumulation of enormously large amounts of data, which facilitates data-driven business innovation, through the deployment of proper big data technologies [66]. Information technology-based firms, such as Amazon, Google, eBay, TikTok, Twitter, etc., constantly store and analyze customer service data, including customer and product details, transaction times and other service-related data. These data sources provide companies with the basis for effective decision making and efficient strategic business plans, to enhance their customer service performance and business innovation [54,67,68]. Research studies have shown that big data technologies are becoming essential for business projects with high technological novelty or radical innovation, especially in a big data environment, as many researchers believe that customers are sources of information and knowledge [69,70]. In [70], four potential key success factors were identified for organizations to integrate big data technologies, so as to accelerate their product innovation processes [71].

A comprehensive investigation has been conducted on the big data technologies that have been applied in the business and financial sectors, including the major methods, technologies and models used for business/data analytics, as shown in Table 1. A number of major big data technologies are listed in Table 1, in accordance with their relevant business and financial activities, as well as the real-life industrial applications or platforms where big data technologies are implemented. Our study discovered that big data technologies have been applied in almost all major activities in the business and financial sectors, as shown in Table 1 [72–79].

Table 1. Big data technologies applied in the business and financial sectors [72–79].

Big Data Technologies	Business/Financial Activities	Industrial Applications/Platforms
Advanced message queuing protocol (AMQP), XMPP, Extract, Transform, Load (ETL)—NoSQL, etc.	Business data acquisition, data cleaning, data pre-processing	• Seon's fraud data detection (Weld ETL tool) • Lenovo's data cleansing (Talend Data Fabric), etc.
Hadoop, Hive, Hydra, Pig, Spark, Mapreduce, Storm, Segmentation (NAD, Bootstrapping), etc.	Data storage, data management, data infrastructure, data migration	• eBay (Apache Spark, Storm, Kafka) • Netflix's media data (AWS s3 & EC2) • Uber (Schemaless, Hadoop), etc.
Collaborative filtering (recommender), linear regression, K-means clustering, apriori association rule, C4.5 (Decision Tree), SVM, etc.	Business analytics, sale prediction, market prediction, financial investment trends analysis	• Walmart (marketing, promotion, personalized customer service) • Uber (Horovod-distributed deep learning framework), etc.
Attribute-based encryption, 3KDEC, storage path encryption, differential privacy, fast anonymization of big data streams, top-down specialization, etc.	Business data privacy, data security, data recovery, big data encryption	• Netflix (released personally identifiable information, removed movie ratings) • IBM-developed ethical framework (for legitimate big data collection), etc.

3. Trends in Big Data Technologies

Big data technologies are now critical to the success of organizations and companies. Numerous big data technologies have been developed and applied to almost every aspect of our daily lives. Online shopping companies, such as Amazon, eBay, Alibaba, Walmart, etc., utilize advanced data mining methods, to provide customers with personalized shopping services and recommendations. Logistics and transportation companies, such as Fedex, DHL, UPS, CSX Transportation, etc., have deployed sensor-based big data technologies and machine-learning methods to improve the efficiency of "last mile delivery".

Big data is a technology with a very broad landscape [80]: it covers various domains and fields, including infrastructure, analytics, applications, data resources, data sources, APIs and open sources. According to the big data technologies statistics in [80], the most influential big data technologies have been adopted in the health sector, particularly in cancer classification, in terms of academic citations [80]. With regard to industrial applications, online shopping and e-commerce are the domains that rely heavily on big data technologies.

In this paper, big data technologies are grouped into four categories: big data acquisition and pre-processing; big data storage and infrastructure; big data analytics; and big data privacy and security. Table 2 provides an overview of current major big data technologies and their trends. Future big data technologies will focus on providing advance solutions, so as to ensure that they can handle more diverse data sources, process heterogeneous data storage, provide real-time analytical solutions and protect the security and privacy of user data more efficiently.

Table 2. An overview of current major big data technologies.

Categories	Current Big Data Technologies	Future Trends
Big data acquisition and pre-processing	• Advanced message queuing protocol (AMQP)—acquisition protocol [75,81]; • Extensible messaging and presence protocol (XMPP)—acquisition protocol [82]; • Java Message Service (JMS)—acquisition protocol [75]; • Extract, Transform, Load (ETL/ELT)—data integration [75]; • Crowdsourcing—data curation [75]; • NoSQL—data structure [75,83]; • Kafka, Flume—platforms [75]; • Crawling—acquisition tool [75,77];	Future big data acquisition and pre-processing technologies should be able to deal efficiently with more unstructured, high dimensional data; several techniques are suggested below: • Blockchain—acquisition protocol, integration and curation [84]; • Data fabric—data acquisition and integration [85]; • Natural Language Processing (NLP) pipelines—data structure and integration [86].
Big data storage and data infrastructure	• Distributed file systems: HDFS, GFS; • Data query analytical tool: Hive QL; Pig Latin; JAQL, etc., [87]. • Storage reduction and optimization: Mapreduce (Hadoop); Spark; Hydra; Storm, etc., [88]; • Data segmentation: (NAD [74], bootstrapping [89]), etc.;	Future trends in big data storage methods will focus on more elastic and cloud-based solutions: • Fog-to-Hybrid and Multicloud [90]; • Data as a Service Model [91]; • Blockchain data storage [92]; • Non-Volatile Memory Express (NVMe) [93] and tensor networks; • Covariance/matrix-based high-dimensional data segmentation and tensor networks [94].
Big data analytics	• Recommendation: collaborative filtering; apriori association rule and FP growth tree, etc., [95,96]. • Unsupervised: K-means clustering; DBSCAN clustering; OPTICS (Ordering points to identify clustering structure); SOM (self-organizing map), etc., [72,97,98]; • Supervised: linear regression; C4.5 (Decision Tree); SVM; CNN; LSTM, etc., [72,99].	The trends in big data analytics are in the following areas: • Natural language processing and sentiment analysis (unstructured data or semi-structured data) [100]; • High-performance, real-time analysis models and virtual reality/Metaverse analytics models [101]; • Data-centric AI analytics and augmented data analytics [102].
Big Data privacy and security	• Data encryption: attribute-based encryption; public key encryption; storage path encryption [78,103]; • Data privacy: privacy-preserving machine learning algorithms [104]; differential privacy [78,105]. • Big data storage privacy and security: Apache Rhino; Sentry; Ranger [106].	The trends in big data privacy and security will mainly focus on cloud- and blockchain-related areas: • Cloud Security—combination of centralized cloud services with fog computing [107]; • Blockchain secure data privacy and security [90,107].

The tensor network is regarded as an efficient solution for future trends in big data storage [108]. The Tensor Train (TT) decomposition, described in [109], can be executed in a simple non-recursive form. A dth-order tensor is defined to be the TT-format, if it satisfies the following format [108]:

$$A(i_1, i_2, \cdots, i_d) = A_1(i_1)A_2(i_2)\cdots A_d(i_d) \tag{4}$$

where $A_{(k)}(i_k) \in R^{r_{k-1}\times r_k}, r_0 = r_d = 1,\ A_{(k)}(i_k)$ is referred to as the core tensor. The TT format is based on a low-rank approximation, through singular value decomposition of auxiliary unfolding matrices [108].

Efficient data segmentation methods have been suggested in various studies [74,93]. A recent study showed that future trends in big data storage are focusing on converting very large datasets to smaller segments without information losses [74]. The Normal Distribution Approximation (NDA) method is adopted for fast large data segmentation, which can be split with minimum data information loss; however, the NDA or the Poisson Distribution Approximation (PDA) methods can only process a one-dimensional data source. Future trends will focus on high-dimensional datasets that take covariance into consideration [94], defining an f ([x_1, x_2 . . . x_i]) function, to judge how close the mean of the subdataset is to the mean of the original dataset. The f function's expression is:

$$([x_1,\ x_2 \ldots x_i]) = (X_1 - x_1)^2 + (X_2 - x_2)^2 + \ldots + (X_i - x_i)^2 \tag{5}$$

where i is the size of the dataset's dimension, [x_1, x_2 . . . x_{i}] is the subdataset's mean, and [X_1, X_2 . . . X_i] is the original dataset's mean. This process involves the calculation of the covariance matrix Σ_{A} (n × n) of the original dataset A and the covariance matrix Σ_{B} (n × n) of the data subset B. A positive definite matrix Σ can be decomposed into $\Sigma = U^T\Lambda\, U$, where U is the upper triangular matrix, and Λ is the diagonal matrix, in which the diagonal elements are non-negative [94]:

$$\sum = U^T\Lambda U = \left[U^T\Lambda^{\frac{1}{2}}\right]\left[\Lambda^{\frac{1}{2}}U\right] = \left[\Lambda^{\frac{1}{2}}U\right]^T\left[\Lambda^{\frac{1}{2}}U\right] \tag{6}$$

and, therefore, the matrix $\Sigma = C^TC$, where $C = \Lambda^{1/2}U$ [94].

The optimized bootstrap algorithms could affect the future trends for big data storage [89]. The bootstrap weights algorithm can be applied to estimate the variance of smooth and non-smooth parameters [89]. The bootstrap weight is defined as:

$$w^*_{ik} = \left\{1 + \left(\frac{n'}{n-1}\right)^{1/2}\left(\frac{nn^*_i}{n'} - 1\right)\right\}\pi_i^{-1}\pi_{k|i}^{-1} \tag{7}$$

where n is the sample size of the primary sample units (psu), and n^*_i denotes the number of times the ith psu is selected in the bootstrap sample; $\widehat{\theta}^*$ is then computed, using the formulae that were used to obtain the original point estimator, with the original weights replaced by the bootstrap weight, w^*_{ik} [94]. The bootstrap variance estimator, $var^*(\widehat{\theta}^*)$, can be applied when all $\widehat{\theta}^*$ are obtained. The Monte Carlo approximation of $var^*(\widehat{\theta}^*)$ is adapted:

$$\widehat{var^*} = \frac{1}{B-1}\sum_{b-1}^{B}\left(\widehat{\theta}^*_b - \widehat{\widehat{\theta}^*}\right) \tag{8}$$

In the future big data environment, unstructured data sources will grow exponentially, and the trends in efficient and real-time data storage solutions will be crucial. The fast-growing Virtual Reality (VR), Augmented Reality (AR), Extended Reality (XR) and Metaverse applications are becoming an increasingly significant issue for big data models. Research studies indicate that the VR and Metaverse will replace the current web-based online businesses, and become the major areas of generating large data volume in the near future [101,110–113]. The following section investigates the current development of the Metaverse and the digital human.

4. Metaverse-Related Technologies and Applications

The definition of the Metaverse in [114] is a virtual space where users can interact with one other, and with their environment, via 3D digital objects and virtual avatars, in a complex manner that mimics the real world, holding things developed using artificial intelligence techniques; therefore, creating digital humans is essential to the development of the Metaverse and other AR/VR/XR applications.

4.1. Digital Human Reconstruction

How to create digital humans has been a much-studied subject recently, due to the rising demand for virtual reality applications, including the Metaverse. One of the core drivers of mathematical progress is the discovery of objects, patterns and ultimately their formulaic representations; in the course of such progress, scientists often need to leverage a variety of tools and data to help them cultivate ideas, propose a conjecture, and eventually prove/disprove with experiments and evidence, where possible. There is no doubt that the evolution of computational methodology has not only changed the way scientists conduct their studies, but has also accelerated the life cycle of scientific research, leading to profound impacts on people's daily lives—including, for example, the early hand-calculated prime number tables used by Gauss (which led to the prime number theorem) [114], the RSA public key algorithm [115] inspired by prime number theory, and our modern blockchain infrastructure.

The introduction of computational methodology has given scientists an understanding of problems previously incomprehensible; however, while previous computational methodologies have proven effective in certain scientific problems or domains, they are not easily generalized to other domains. Big data technologies, especially the field of deep learning that has emerged in recent years, offer a range of techniques capable of effectively detecting patterns in data, and are increasingly proving their utility in scientific disciplines. A specific case of virtual human reconstruction in the Metaverse will serve as an example, to illustrate how deep learning can be used to solve mathematical problems in practical settings.

Virtual human reconstruction is one of the essential tasks in various Metaverse applications: it aims to utilize sensory data to recover the three-dimensional geometry and appearance of humans, achieving accurate photorealistic reconstructions, and ultimately producing compact 3D representations that can be ported to a variety of devices. This problem involves many practical facets that require sophisticated engineering; however, its core challenges lie in deep learning modeling and mathematical optimization, as shown in Figure 6.

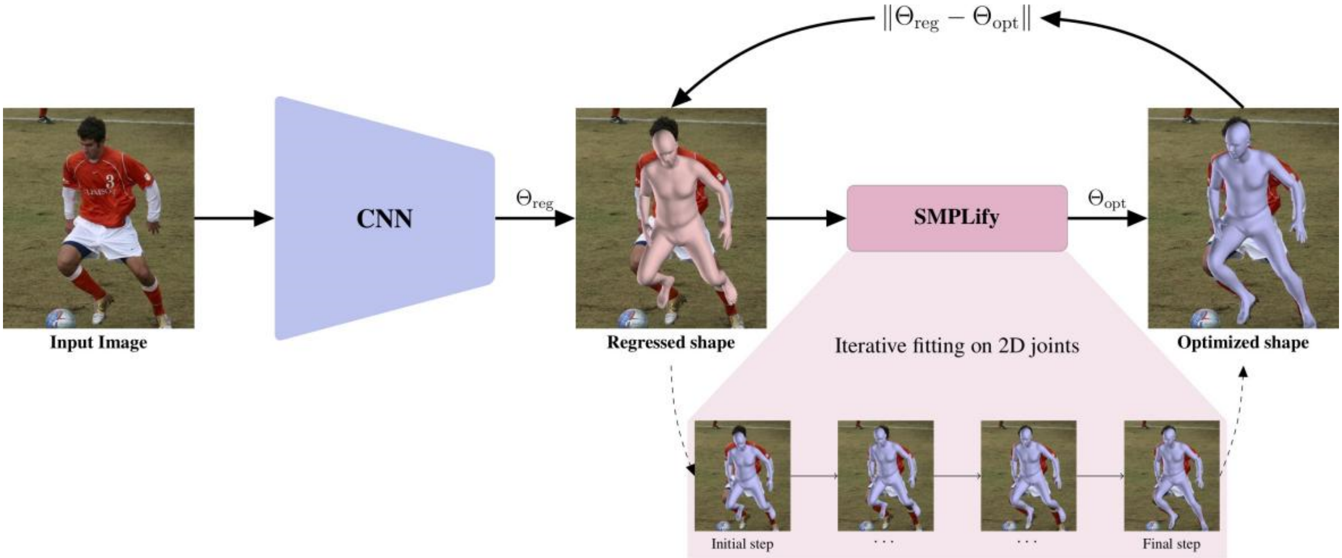

Figure 6. A hybrid approach of regression-based and optimization-based paradigms (courtesy of Kolotouros et al. [116]): an iterative optimization routine is embedded into a neural network training loop, leading to a self-improving loop. Better fits help the network train better, while better initial estimates from the network help the optimization routine converge to better fits.

Various techniques have been applied to recreate human models in the Metaverse. Many studies start from simple image-based 2D feature detection, such as key points [117], silhouettes [118] and limb segments [119]. It seems that simple movements can be represented relatively clearly by two-dimensional contents; however, it is becoming clear that complex human behaviors, which often occur in practical settings, do not fit the simple assumptions imposed by two-dimensional models, and that more descriptive models with finer granularity are desirable; consequently, more studies [120–122] have turned to exploring more complex human pose modeling in three dimensions. Recently, researchers have noticed that body shapes, contacts, gestures and expressions which directly interact with the world are much easier to measure and evaluate; consequently, the focus of researchers has shifted towards three-dimensional mesh recovery of the human body [123,124]. Human body modeling is then further extended by face and hands support [125–128]. Meanwhile, similar techniques have also facilitated downstream tasks, such as clothed human reconstruction [129–131], volume rendering [132], virtual try-on [133], the computer-assistant system [134] and many more Metaverse applications. There are two common paradigms for dealing with virtual human reconstruction: the optimization-based paradigm (described in Section 4.3) and the regression-based paradigm (described in Section 4.4).

Although these two paradigms may have different advantages/disadvantages, and address different aspects, both paradigms can share similar human body modeling techniques. Figure 7 shows an interesting possible way of integrating both paradigms into one coherent framework. The next section will review the existing approaches, in terms of human body modeling.

Figure 7. A virtual reality shop developed by Unity3D for future integration into the Metaverse.

4.2. Review of Human Body Modeling

Early human body modeling started with the study of articulated geometric primitives, including line segments [135], cylinders [136], planar rectangles [137] and ellipsoids [138]. As three-dimensional full-body scanners became accessible, more detailed measurements of body surfaces could be accurately recorded, such as the CAESAR (Civilian American and European Surface Anthropometry Resource) [139] dataset. The availability of large amounts of body scan data has given rise to a powerful representation: the statistical body model, which factors body deformations into identity-dependent and pose-dependent components. Among the statistical body models, SCAPE [140], SMPL [141], SMPL-X [126], SMPL+H [142], 3DMM [143] and STAR [144] are popular ones, which are not only capable of effectively modeling both shape and pose deformations, but are also highly compatible with existing graphics rendering engines, benefiting from the explicit mesh model. This family of explicit approaches first learns shape deformations through principal component analysis of body scans, and then combines them with skeletal pose-driven deformations (so-called linear blend skinning in traditional skeletal animation), to construct a shape-and-pose parametric human body model. Despite the popularity of explicit approaches, they still have their limitations: firstly, global blend shapes may capture spurious long-range correlations [144], resulting in non-local deformation artifacts; secondly, correlations between body shape and pose-dependent shape deformation may be ignored; furthermore, due to the linear nature of principal component analysis, it can be difficult to reproduce the highly nonlinear deformations of body soft tissue.

In order to overcome the limitations of explicit approaches, instead of explicitly defining the human body as mesh vertices and edges or other elements, implicit approaches try to define surfaces as level sets of continuous functions. Due to these continuous properties, this implicit representation has a better chance of being elegantly optimized and integrated with deep learning frameworks: it is continuous across the spatial domain, and thus theoretically has infinite resolution, and it can easily handle highly nonlinear deformations, and even topological changes, which are not possible with explicit approaches. Study [145,146] estimated implicit surface functions, by aligning image pixels with the global three-dimensional shape or texture of the photographed object, and then using a dedicated multi-level network to refine the resulting geometry. The flexibility of implicit approaches enabled it to handle intricate surfaces and topological changes with ease, but there was one drawback, which was that topologically distinct human representations can exist across time: in other words, implicit human representations may not be topologically consistent in time.

4.3. Optimization-Based Paradigm

In this paradigm, the human body model is explicitly optimized, by minimizing an objective function that fits the model to the observations in an iterative manner. The objective function typically consists of two parts: (1) the data term is a measure of the alignment between the extracted observation features and the transformed human body features; (2) the regularization term is added, to constrain the convergence that preserves a physically plausible body model. In earlier work, the silhouette feature played a crucial role in fitting the body model to the image, as it was used to penalize pixels in non-overlapping regions [147,148].

With the emergence of deep learning, many studies have utilized it to calibrate the optimization initial conditions. SMPLify [123] adopts off-the-shelf neural networks [149] to detect two-dimensional key points, and then iteratively fits a SMPL model, to detect the key points of an unconstrained image. While SMPLify produces relatively well-aligned results, sparse key points do not offer sufficient constraints for body shape optimization. To improve geometric details, [150–152] combined key points, silhouettes and part segments, to further constrain the optimization process. Moreover, [153,154] have shown that deep learning techniques can learn local landscapes and decent directions of optimization from training data, and then use them to guide the gradient-based optimization process: in this way, traditional problem-independent optimization schemes can be endowed with the ability to adaptively learn problem-specific convergence schemes. Image-based key point regression was performed by [155,156], to obtain three-dimensional body key points, then solve the inverse kinematics based on the key points and the skeletal structure, so as to calculate the accurate joint rotations, ultimately estimating the parameters of a SMPL model.

Although the optimization-based paradigm can faithfully reconstruct the human body when high quality data is available, it performs poorly in situations where data is scarce and useful information is latent; furthermore, as the optimization-based paradigm intrinsically tries to solve complex non-convex optimization problems in high-dimensional spaces, its outcomes are susceptible to initialization and prone to falling into spurious local minima.

4.4. Regression-Based Paradigm

Alternatively, the regression-based paradigm exploits the powerful learning and approximation capabilities of neural networks, to recover model parameters directly from sensory data. To achieve better performance, researchers have explored a wide variety of network architectures and regression objectives—for example, [125] was one of the pioneering efforts to incorporate the SMPL model into an end-to-end network architecture that minimized the reprojection errors between manually annotated and estimated key points. An end-to-end adversarial learning framework was proposed by [124], which used a discriminator to supervise the training process, so as to exclude anthropometrically implausible or self-intersecting body structures. A top-down framework was proposed by [157], to simultaneously regress SMPL parameters of multiple people in a coherent manner, where depth ordering was consistent, and no interpenetration occurred among reconstructed people. Instead of regressing the SMPL parameters, [158] opted to directly regress the mesh vertices using a Graph Convolutional network, thus allowing the template mesh structure to be explicitly encoded within the network, easily exploiting the mesh spatial locality. Inspired by [124], VIBE [159] went a step further, to estimate dynamic motion sequence from videos. By replacing the regression network with a temporal generative network, and changing the three-dimensional supervision dataset to a motion capture dataset, AMASS [160], VIBE empowered an adversarial learning framework with temporal information, enabling motion sequence estimation as a whole.

To leverage expressive human models and paired data, [127,161,162] adopted a divide-and-conquer strategy, by breaking down the human reconstruction problem into part-specific estimation subproblems, where body, hand and face estimates were performed using the respective part-specific models. The final expressive model was obtained by assembling the individual results of the subproblems into the corresponding body template layers. ExPose [127] directly regressed hands, face and body parameters in the SMPL-X format, and utilized body-driven attention to localize the face and hands regions for refinement, using part-specific knowledge learned from existing face- and hand-only datasets. A real-time method was introduced by [163], to capture body, hands and face with competitive accuracy, by exploiting correlations between body and hands. Pose2Pose [164] extracted joint-specific local and global features, to train a graph convolutional neural network, and regress body/hand joint rotations from it. PIXIE [161] first fused the features from body, face and hand experts, according to their part-specific confidences, and then fed these features into the part-specific networks, for robust regression.

4.5. Technologies in AR/VR/XR Platforms and the Metaverse: Future Trends

In our opinion, AR/VR/XR applications will undoubtedly, in the near future, become the ultimate customer service platforms. In other words, AR/VR/XR applications will at least become the dominant platforms, if they do not completely wipe out the current mobile and computer platforms. Consequently, a big data surge will very soon occur in the virtual world. The Metaverse is likely to be the front platform to face the data surge challenge, due to its rapid growth in recent years. The following figure shows our recently developed VR-based shopping platform.

The authors observed that two extreme situations would occur in the Metaverse, while conducting user recommendation and data analysis: (1) *The cold start problem*. This situation often occurs when too little data is available for data analysis, due to the VR platforms being new to users, and to not much information having been generated and accumulated for analysis, a common situation in the big data environment, when new platforms are released for users; (2) *The virtual data explosion problem*. This situation occurs when the Metaverse or VR platforms generate too much data, including user interaction data, wearable sensor data, eye tracking data, location trajectory data, brain EEG data, and business transaction data. Figure 8 shows the data sources of the Metaverse and its architecture [165], which indicates that the Metaverse consists of various data sources from physical, social and digital worlds.

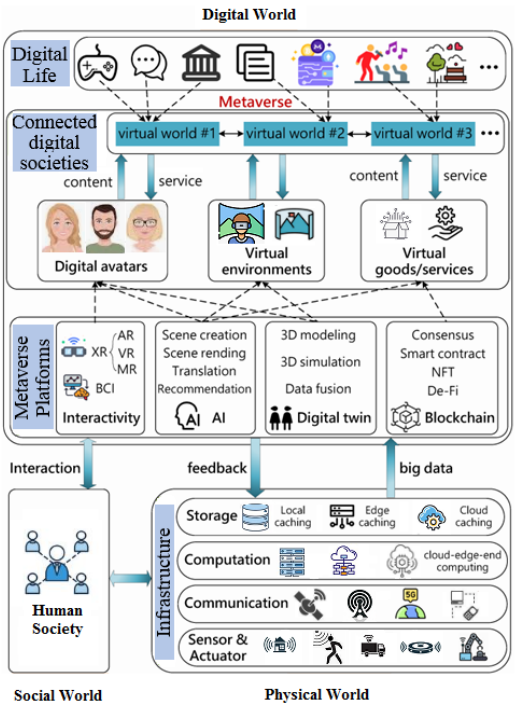

Figure 8. Metaverse architecture of integrated social, physical and digital worlds, modified based on [165]. The social world mainly consists of human communities.

Several methods have been suggested for solving the abovementioned problems. In [166], a position-based VR online shopping recommendation system was developed, to solve the cold start problem in VR platforms. In such a system, the cold start problem is tackled by analyzing new users' interaction and behaviors within the virtual world. For instance, the position-based VR online shopping system acquires new users' trajectories in the virtual world, and conducts analysis based on their movements, to generate user recommendations, as shown in Figure 9.

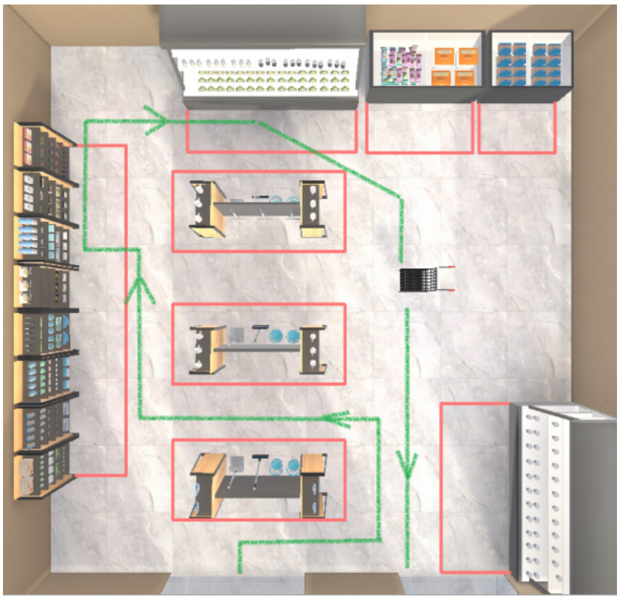

Figure 9. Position-based analysis for VR shopping recommendation (green line is user trajectory).

Future trends in solving the cold start problem in the Metaverse will further utilize users' behavior and sentiment data, including user eye tracking data, user movement trajectory, wearable user device data, and user sentiment data. In particular, human brain data analysis will likely become an essential technology for user analysis in VR platforms, such as the Metaverse.

The cold start problem is not a persistent problem in VR platforms, as it can be solved automatically when data accumulation reaches a certain quantity, whereas the virtual data explosion problem is a persistent challenge to VR platforms like the Metaverse. The wide range of data sources in the Metaverse will grow exponentially, due to its digitization in nature. Some research studies have suggested adopting the Data as a Service (DaaS) framework [91], as the solution to the data explosion problem in the digital world, including the Metaverse. Several other solutions, including tensor networks and sentiment analysis, have been proposed, to solve this problem. The future trends of technical development in the Metaverse and other VR platforms can be summarized as follows:

(1) **Digital human reconstruction** is becoming a crucial area for the Metaverse and other VR platforms: this is a core technology that can accelerate the development of the Metaverse, so as to truly realize human–machine interaction in virtual worlds, as mentioned in Sections 4.1–4.3;

(2) **Digital Twin**-related methods are the foundation for creating digital worlds that can mimic the physical world. The digital twin is defined as the effortless integration of data between a physical and virtual environment, in either direction [167]. VR-developing tools, such as Unreal Engine, Unity, 3DS Max & Maya, SketchUp, etc., will be the major developer's toolkits for digital twin models in the coming decades. The future trends in digital twin will focus on the following: enabling a conformance relationship between digital twin and the real world; digital world autonomy, runtime self-adaptation and self-management; and integration and cooperation, to achieve

common goals or provide services [168]. A number of digital twin applications have been developed, based on Microsoft Kinect sensors and the Oculus VR headset.

(3) **Brain–Computer Interface** (BCI) technology will become a very important area for the Metaverse and for VR platforms. Previous research indicates that non-invasive BCI technology has been applied extensively in various areas in recent years, because of its minimal potential risks and time precision [169]. Figure 10 shows the high-performance EEG BCI method (left), and EEG BCI experiments (right) [169,170].

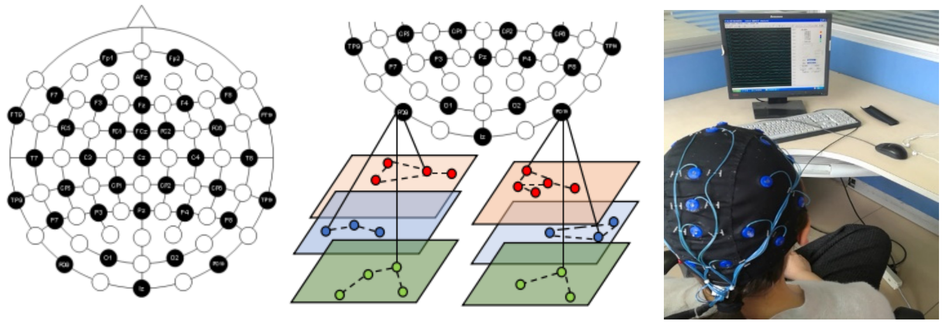

Figure 10. Segmented EEG time window (**left**), source: [169]; EEG experiment (**right**), source: [170].

The NDA/PDA-based methods are adopted, to enhance EEG data analytical efficiency, in order to accommodate the real-time interaction in the Metaverse and VR platforms [74]. The definition for the NDA method is as follows: if S [a, b] $\subseteq A$ [1, k], if $x \in$ [a, b] satisfies:

$$f(A(x), \mu, \sigma) = \frac{1}{\sigma}\Phi(\frac{A(x)-\mu}{\sigma}) \xrightarrow[\substack{a \leq x \leq b \\ a \neq b}]{} S[a,b] \subseteq ND \tag{9}$$

$$\Phi(S) \geq (1-m_r) \times \frac{1}{\sigma\sqrt{2\pi}}\int_a^b \exp(-\frac{c^2}{2})dx \tag{10}$$

where m_r is the adjusting parameter, and S [a, b] is an NDA set. The ND-based method derives the data values using *ksdensity* function, to generate a probability distribution [170]. The definition for the PDA method is as follows: the PDA model takes one of the calculated σ and λ values as $\lambda \times t$, as indicated in the following equations, 11 and 12. Assuming the original data set has σ, then Mean (λ) is the event rate. If Mean (λ) − λ = Δ, then $\lambda \times$ t is lying between Mean (λ) and λ. With ∣y − $\lambda \times t$∣ = a, $a^{1/2}$+a = Δ is satisfied.

$$P(k\ events\ in\ fix\ time) = e^{-\lambda}\frac{\lambda^k}{k!} \tag{11}$$

$$P(N(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \tag{12}$$

where *N(t)* is the sample data in the *t* time window. The Gamma function is utilized in the PDA method for processing complex numbers, which is expressed in (13) below [171]:

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx \tag{13}$$

The Δ parameter is used to regulate the size of the sample data sets, to get the nearest λ and σ values. The Δ parameter in the PDA plays the same role that it plays in the NDA method. The PDA model employs a PDA benchmark point selection method [169–171].

(1) **Blockchain** technology is an efficient and secure solution for digital worlds, such as the Metaverse. In the blockchain model, a new transaction can be verified and added to existing records, i.e., blocks, through linking the new transaction to previous ones, by cryptographic hash operation [172]. Each block contains a cryptographic

hash of the previous block, a timestamp, and transaction data [173]. The main characteristics of blockchain technology are that it is secure, decentralized, digitized, collaborative and immutable: these characteristics make blockchain technology a perfect solution for digital virtual worlds, such as the Metaverse. Currently, the most successful security technology for blockchain employs the Public Key Infrastructure (PKI)-based blockchain methods [174]. Researchers in the field have started to search for more efficient solutions. The future trends in blockchain technology development in the Metaverse intend to focus on more autonomous, intelligent and scalable models, such as intelligence-agent-based blockchain [175], Self-Sovereign Identity (SSI) blockchain [176], non-fungible tokens (NFTs) [177] and bio-identity-based blockchain.

(2) **Artificial intelligence (AI)** is a discipline essential to almost all areas in our modern world, particularly for future virtual worlds such as the Metaverse. AI can accelerate analytical efficiency, enhance security and privacy, improve interoperability, and provide better solutions for human–machine interaction and collaboration. The increase in applications of Natural Language Processing (NLP), sentiment analysis and brain informatics technologies to digital worlds is stimulating the development of AI in these areas. The successful stories of AI implementation in image recognition, voice recognition, human–machine interaction and intuition, reveal the promising future of AI in the Metaverse and other virtual worlds. A recent survey showed that a majority of studies had focused on exploring efficient integration and collaboration between Edge AI architecture and the Metaverse [178].

5. Discussion

Virtual platforms, such as the Metaverse, and big data technologies are becoming a part of our daily lives. The development of these two different but closely related technologies has accelerated in recent years. In this section, the authors would like to discuss the role of big data in the development of the Metaverse, by reviewing a chronicle of the Metaverse. This section provides the literature review methods used in this survey.

5.1. A Chronicle of the Metaverse and the Role of Big Data in the Metaverse

Figure 11 demonstrates how the Metaverse and its related technologies, which include big data, have evolved and developed [178].

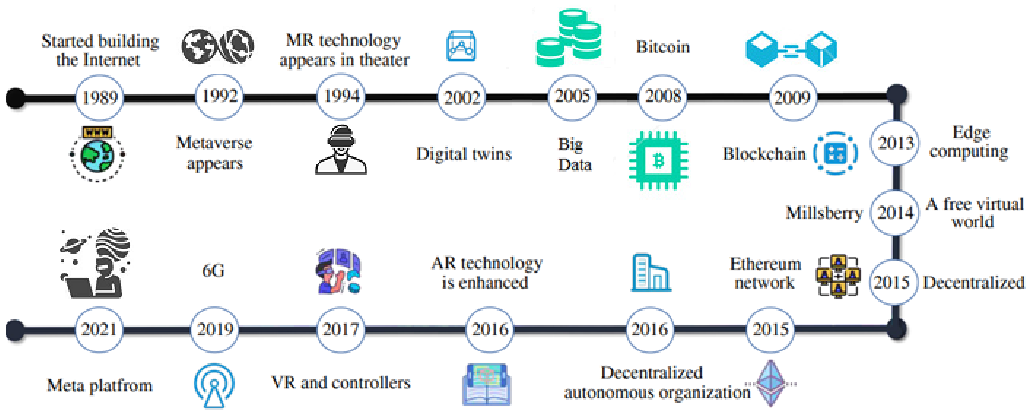

Figure 11. A chronicle of the Metaverse and its related techniques, modified based on [178].

Data sources in the Metaverse and other virtual platforms are growing exponentially; therefore, big data technologies are crucial for the Metaverse, if it is to efficiently manage its digital world, and provide users with real-time analytical services. Big data technologies are fundamental tools for rendering virtual platforms, such as the Metaverse, feasible for users. In other words, big data is a fundamental component in the Metaverse; and the Metaverse accelerates the development of big data technologies; however, big data is not only crucial in the virtual world—it is also an important component of our real physical world, as evidenced in various areas, including the fields indicated in Section 1. Figure 12 shows the relationship between big data and the Metaverse.

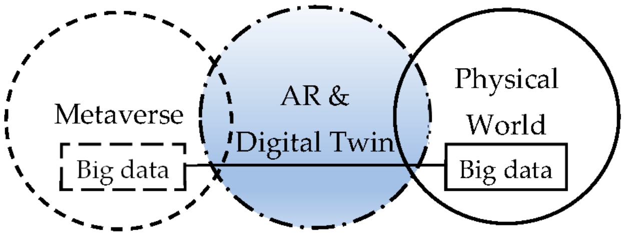

Figure 12. Big data plays a key component in both the physical world and virtual worlds. The Metaverse is a virtual world parallel to the real physical world: the two are sometimes connected by augmented reality and digital twin.

The current definitions of the Metaverse vary according to different studies; however, many researchers share a common view that the Metaverse is imitating our physical world. In this survey, the authors believe that future virtual worlds, including the Metaverse, will develop to be totally different world from our physical world: these virtual worlds will go beyond our current social structure and civil life. Table 3 shows the example applications of the Metaverse and big data in several key sectors.

Table 3. A brief review of example applications of big data and the Metaverse in major sectors.

Sectors	Big Data	Metaverse
Healthcare	• Real-time big data analytical models (Health-CPS); • Data as a Service e-health systems, etc., [179].	• Metaverse hospital (Thumbay, Davita); • Interactive diagnosis platforms, etc., [180].
Finance and Economy	• Big data finance and business analytics (Splashback); • Online business decision support, etc., [72–79,181].	• Metaverse banks (Onyx, ZELF); • NFTs, Bitcoins, VR-funds, etc., [182,183].
Education	• Learning performance analysis and customization [184]; • Education data warehouse, BD curriculum, etc., [185].	• Metaversity (Novartis, King's InterHigh); • Immersive realistic learning scene [186].
Entertainment and Social	• User behavior and opinion analysis, social trends [187]; • Game data monitoring, sentiment analysis [188].	• Metaverse games (Roblox, Sandbox) [189]; • Virtual social (Meta, Altspace VR) [190].

5.2. Literature Review Methods

This research adopted comprehensive and systematic literature review methods, which consisted of the background and concepts review, survey data retrieval tool selection, and survey data extraction and analysis, summarizing future trends based on analysis and comparison.

This research followed the principles of the literature review outlined in [191], and conducted three types of review methods in the background and concepts review process, including narrative review, systematic review and integrative review [192]. The review of big data and Metaverse applications in different sectors was based on the narrative and systematic review methods. The review of the trends in big data and the Metaverse adopted the integrative review method. The purpose of the big data and Metaverse applications review was to explore the latest developments in big data and Metaverse technologies in the industrial sectors; therefore, a large number of literature and online resources were obtained for narrative review in this process. The purpose of investigating the trends in big data and Metaverse development was to provide readers with a future perspective on big data and Metaverse technologies, and to outline the emerging new methods and the potential paradigm shift. The survey data sources and retrieval tools used in this paper are listed in Table 4.

Table 4. Literature review methodologies.

Review Content	Review Approach	Retrieval Tools	Data Sources
Big data applications in different sectors (Section 1)	Narrative and systematic review	• Digital library • Workshop discussion • Online survey	• IEEE/ACM online • IDP2022 forum • WJX survey
Metaverse technologies and applications (Sections 4.1–4.4)	Narrative and systematic review	• Digital library • Online search engine • Workshop discussion	• IEEE/ACM online • Google search • IDP2022 forum
Future trends in big data and Metaverse technologies	Integrative review	• Online search engine • Digital library • CiteSpace	• Google search • IEEE/ACM online • Web of Science
Correlation of big data and Metaverse and future development	Systematics and integrative review	• Digital library • CiteSpace • Online search engine • Workshop discussion	• IEEE/ACM online • Web of Science • Google search • IDP 2022 forum

The analytical methods of the literature review consisted of CiteSpace analytics [193], Google Scholar [194], Semantic Scholar [195] and the IEEE digital library: these tools provided the authors with the literature chronicle review, citation and impact factor analyses, and the trends analysis. Figure 13 shows the CiteSpace tool that was used in our previous survey [196] for BCI analysis associated with this research.

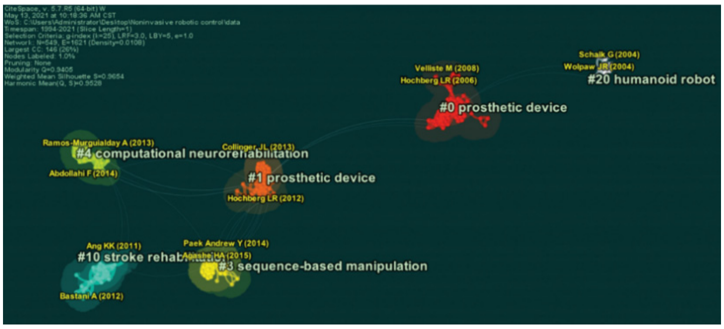

Figure 13. BCI literature review analysis based on CiteSpace [196].

6. Conclusions

This survey provides a comprehensive investigation of the development of big data technologies and virtual platforms, i.e., the Metaverse. Undoubtedly, the Metaverse and big data technologies will become the most influential areas in the following decades. Big data is an essential component of the Metaverse, and it also plays a crucial role in the parallel real world: these two areas will eventually merge into the virtual world, and advance mutual development.

This research survey grouped the big data technologies into four categories: big data acquisition and pre-processing; big data storage and data infrastructure; big data analytics; and big data privacy and security. The authors further investigated the future trends in each category, and obtained the following findings: (1) future big data acquisition and pre-processing technologies will aim to deal efficiently with more unstructured, high-dimensional data. Among various data acquisition methods, blockchain, data fabric and NLP pipelines demonstrate their potential for future big data acquisition; (2) future trends in big data storage methods intend to focus on more elastic and cloud-based solutions. Fog-to-Hybrid and multicloud, Data as a Service model, blockchain data storage, and high-dimensional data segmentation and tensor networks are the major future trends for big data storage; (3) the trends in big data analytics will focus on natural language processing, sentiment analysis, real-time analytical models and virtual reality/Metaverse analytics models, and data-centric AI analytics; (4) the trends in big data privacy and security will mainly focus on cloud, blockchain-related areas, such as cloud security (combination of centralized cloud services with fog computing), blockchain secure data privacy and security.

The Metaverse and other virtual platforms have grown rapidly in recent years. The PwC Co. predicts that VR and AR platforms will boost global GDP by USD 1.5 trillion by 2030 [197]. To date, applications of the Metaverse have included online shopping, virtual social media, video games, virtual tours, and online museums and arts [111,112,198]. Many large technology companies have announced plans to launch their Metaverse products, such as Facebook Horizon, Nvidia Omniverse, and Amazon Metaverse. The future trends in technical development in the Metaverse and other VR platforms can be grouped into five main areas: digital human; digital twin; brain–computer interface (BCI), blockchain and artificial intelligence. Notably, brain–computer interface technologies have become increasingly important to Metaverse development in recent years, as immersive interactions provided by BCI can enhance user experience [196,199–202].

In summary, our society is becoming more digitized and virtualized. A virtual world, e.g., the Metaverse, can bring our society some overwhelming benefits, including convenience, energy saving, time efficiency, creativity, and being environmentally friendly. Big data technologies, as a fundamental component of the Metaverse, offer methods and algorithms to solve the data explosion and analytical bottleneck problems in the digital world.

The Metaverse is currently imitating our physical real world as a parallel virtual world; however, the Metaverse will become very different from our physical world. The authors believe that virtual worlds, with the help of big data and AI technologies, will become more advanced than our physical world, in many aspects. Big data is undoubtedly one of the most crucial areas for virtual worlds, such as the Metaverse, because virtual worlds are digital data-based, and will accelerate data explosion; therefore, it is crucial to understand the development of big data and of the Metaverse, respectively, and then to find out the common areas in which they would facilitate each other, as shown in Figure 12. Future research will further identify the research problems in the Metaverse and other virtual worlds, in terms of dealing with big data issues in complete virtual or hybrid virtual environments.

Author Contributions: Conceptualization, H.Z.; methodology, H.Z., S.L. and Y.L.; software, H.Z.; validation, H.Z., S.L. and Y.L.; formal analysis, H.Z.; investigation, H.Z.; resources, H.Z., S.L., X.Y. and H.L.; data curation, H.Z.; writing—original draft preparation, H.Z., S.L. and Y.L.; writing—review and editing, H.Z., S.L. and Y.L.; visualization, H.Z. and Y.L.; project administration, H.Z., H.L. and X.Y.; funding acquisition, H.Z., H.L. and X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the Ningbo Natural Science Fund Plan (Grant No. 2022J164), the Humanity and Social Science Foundation of the Ministry of Education of China (21A13022003), the Ningbo Public Welfare Science and Technology Plan (Grant No. 2021S093), the Zhejiang Provincial Natural Science Fund (LY19 F030010), the Zhejiang Provincial Philosophy and Social Sciences Planning Project (20NDJC216YB), the National Natural Science Foundation of China Grant (No. 72071049), the Zhejiang Provincial Philosophy and Social Sciences Planning Project

(22JCXK08Z), the Zhejiang Provincial Natural Science Fund (LZ21A010001), and the Ningbo Key Technology R&D Project (2022Z167).

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the members in the center for SCDM, NIT, Zhejiang University for providing support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Abu-Salih, B.; Wongthongtham, P.; Zhu, D.; Chan, K.Y.; Rudra, A. Introduction to Big Data Technology. In *Social Big Data Analytics*; Springer: New York, NY, USA, 2021; pp. 15–59. [CrossRef]
2. Gutierrez, D. Big Data Business Impact: Achieving Business Results through Innovation and Disruption. 5th Annual Big Data Executive Survey for 2017. Available online: https://www.businesswire.com/news/home/20170109005058/en/NewVantage-Partners-Releases-5th-Annual-Big-Data-Executive-Survey-for-2017 (accessed on 30 November 2022).
3. Chamorro-Premuzic, T. How the Web Distorts Reality and Impairs Our Judgment Skills. *The Guardian*. 13 May 2014. Available online: https://www.theguardian.com/media-network/media-network-blog/2014/may/13/internet-confirmation-bias (accessed on 30 November 2022).
4. Ning, H.; Wang, H.; Lin, Y.; Wang, W.; Dhelim, S.; Farha, F.; Ding, J.; Daneshmand, M. A Survey on Metaverse: The State-of-the-art, Technologies, Applications, and Challenges. *arXiv* **2021**, arXiv:2111.09673.
5. Lin, R.; Ye, Z.; Wang, H.; Wu, B. Chronic Diseases and Health Monitoring Big Data: A Survey. *IEEE Rev. Biomed. Eng.* **2018**, *11*, 275–288. [CrossRef]
6. Shashkina, V. Why Your Medical Organization Can no Longer Do without a Healthcare Data Warehouse. *Itrex Archive*. 3 December 2021. Available online: https://medium.datadriveninvestor.com/why-your-medical-organization-can-no-longer-do-without-a-healthcare-data-warehouse-774cd374e0a (accessed on 30 November 2022).
7. Chawla, N.V.; Davis, D. Bringing Big Data to Personalized Healthcare: A Patient-Centered Framework. *J. Gen. Intern. Med.* **2013**, *28*, 660–665. [CrossRef] [PubMed]
8. Yao, Q.; Tian, Y.; Li, P.-F.; Tian, L.-L.; Qian, Y.-M.; Li, J.-S. Design and Development of a Medical Big Data Processing System Based on Hadoop. *J. Med. Syst.* **2015**, 39. [CrossRef] [PubMed]
9. Zhang, Y.; Qiu, M.; Tsai, C.-W.; Hassan, M.M.; Alamri, A. Health-CPS: Healthcare Cyber-Physical System Assisted by Cloud and Big Data. *IEEE Syst. J.* **2015**, *11*, 88–95. [CrossRef]
10. Chaudhuri, S.; Dayal, U.; Narasayya, V. An overview of business intelligence technology. *Commun. ACM* **2011**, *54*, 88–98. [CrossRef]
11. Rathore, M.M.; Ahmad, A.; Paul, A.; Wan, J.; Zhang, D. Real-time Medical Emergency Response System: Exploiting IoT and Big Data for Public Health. *J. Med. Syst.* **2016**, *40*, 283. [CrossRef]
12. Poon, C.C.Y.; Lo, B.P.L.; Yuce, M.R.; Alomainy, A.; Hao, Y. Body Sensor Networks: In the Era of Big Data and Beyond. *IEEE Rev. Biomed. Eng.* **2015**, *8*, 4–16. [CrossRef]
13. Bussmann, J.B.J.; Martens, W.L.J.; Tulen, J.H.M.; Schasfoort, F.C.; Berg-Emons, H.J.G.V.D.; Stam, H.J. Measuring daily behavior using ambulatory accelerometry: The Activity Monitor. *Behav. Res. Methods Instrum. Comput.* **2001**, *33*, 349–356. [CrossRef]
14. Malasri, K.; Wang, L. Addressing security in medical sensor networks. In Proceedings of the HealthNet'07: 1st ACM SIGMOBILE International Workshop on Systems and Networking Support for Healthcare and Assisted Living Environments, New York, NY, USA, 11 June 2007; pp. 7–12. [CrossRef]
15. Lavigne, T.; Mbarek, B.; Pitner, T. A Real Time Healthcare Tracking System based on Blockchain Application. In Proceedings of the 2021 IEEE/ACS 18th International Conference on Computer Systems and Applications, Tangier, Marocco, 8–11 November 2021; pp. 1–8.
16. Lee, S.; Gandla, S.; Naqi, M.; Jung, U.; Youn, H.; Pyun, D.; Rhee, Y.; Kang, S.; Kwon, H.-J.; Kim, H.; et al. All-Day Mobile Healthcare Monitoring System Based on Heterogeneous Stretchable Sensors for Medical Emergency. *IEEE Trans. Ind. Electron.* **2019**, *67*, 8808–8816. [CrossRef]
17. Chakraborty, C.; Kishor, A. Real-Time Cloud-Based Patient-Centric Monitoring Using Computational Health Systems. *IEEE Trans. Comput. Soc. Syst.* **2022**, *9*, 1613–1623. [CrossRef]
18. Jiang, Y.; Sha, L.; Rahmaniheris, M.; Wan, B.; Hosseini, M.; Tan, P.; Berlin, R.B. Sepsis Patient Detection and Monitor Based on Auto-BN. *J. Med. Syst.* **2016**, *40*, 111. [CrossRef] [PubMed]
19. Dillinger, P.; Vogelbruch, J.F.; Leinen, J.; Suslov, S.; Patzak, R.; Winkler, H.; Schwan, K. FPGA-Based Real-Time Image Segmentation for Medical Systems and Data Processing. *IEEE Trans. Nucl. Sci.* **2006**, *53*, 2097–2101. [CrossRef]
20. He, D.; Chen, C.; Chan, S.; Bu, J.; Vasilakos, A.V. ReTrust: Attack-Resistant and Lightweight Trust Management for Medical Sensor Networks. *IEEE Trans. Inf. Technol. Biomed.* **2012**, *16*, 623–632. [CrossRef]
21. Zwolenski, M.; Weatherill, L. The Digital Universe Rich Data and the Increasing Value of the Internet of Things. *J. Telecommun. Digit. Econ.* **2014**, *2*, 1–9. [CrossRef]

22. Mønsted, T. Achieving veracity: A study of the development and use of an information system for data analysis in preventive healthcare. *Health Inform. J.* **2017**, *25*, 491–499. [CrossRef]
23. Naeini, E.K.; Azimi, I.; Rahmani, A.M.; Liljeberg, P.; Dutt, N. A Real-time PPG Quality Assessment Approach for Healthcare Internet-of-Things. *Procedia Comput. Sci.* **2019**, *151*, 551–558. [CrossRef]
24. Naeini, E.K.; Shahhosseini, S.; Kanduri, A.; Liljeberg, P.; Rahmani, A.M.; Dutt, N. AMSER: Adaptive Multimodal Sensing for Energy Efficient and Resilient eHealth Systems. In Proceedings of the 2022 Design, Automation & Test in Europe Conference & Exhibition (DATE), Antwerp, Belgium, 14–23 March 2022; pp. 1455–1460. [CrossRef]
25. Muhammad, G.; Alshehri, F.; Karray, F.; El Saddik, A.; Alsulaiman, M.; Falk, T.H. A comprehensive survey on multimodal medical signals fusion for smart healthcare systems. *Inf. Fusion* **2021**, *76*, 355–375. [CrossRef]
26. Deshpande, P.; Rasin, A.; Tchoua, R.; Furst, J.; Raicu, D.; Antani, S. Enhancing Recall Using Data Cleaning for Biomedical Big Data. In Proceedings of the IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), Rochester, MN, USA, 28–30 July 2020; pp. 265–270.
27. Zhang, J.; Wang, F.-Y.; Wang, K.; Lin, W.-H.; Xu, X.; Chen, C. Data-Driven Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 1624–1639. [CrossRef]
28. Shi, Q.; Abdel-Aty, M. Big data applications in real-time traffic operation and safety monitoring and improvement on urban express ways. *Transp. Res. C Emerg. Technol.* **2015**, *58*, 380–394. [CrossRef]
29. Zhu, L.; Yu, F.R.; Wang, Y.; Ning, B.; Tang, T. Big Data Analytics in Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 383–398. [CrossRef]
30. Mohamed, N.; Al-Jaroodi, J. Real-time big data analytics: Applications and challenges. In Proceedings of the 2014 International Conference on High Performance Computing & Simulation (HPCS), Bologna, Italy, 21–25 July 2014; pp. 305–310. [CrossRef]
31. Lin, X.; Wang, P.; Wu, B. Log analysis in cloud computing environment with Hadoop and Spark. In Proceedings of the 2013 5th IEEE International Conference on Broadband Network & Multimedia Technology, Bristol, UK, 17–19 November 2013; pp. 273–276. [CrossRef]
32. Shan, Z.; Zhao, D.; Xia, Y. Urban road traffic speed estimation for missing probe vehicle data based on multiple linear regression model. In Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, 6–9 October 2013; pp. 118–123.
33. Zenina, N.; Borisov, A. Regression Analysis for Transport Trip Generation Evaluation. *Inf. Technol. Manag. Sci.* **2013**, *16*, 89–94. [CrossRef]
34. Abellán, J.; López, G.; de Oña, J. Analysis of traffic accident severity using Decision Rules via Decision Trees. *Expert Syst. Appl.* **2013**, *40*, 6047–6054. [CrossRef]
35. Xie, C.; Lu, J.; Parkany, E. Work Travel Mode Choice Modeling with Data Mining: Decision Trees and Neural Networks. *Transp. Res. Rec. J. Transp. Res. Board* **2003**, *1854*, 50–61. [CrossRef]
36. Silva, C.; Saraee, M. Predicting Road Traffic Accident Severity using Decision Trees and Time-Series Calendar Heatmaps. In Proceedings of the 2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET), Penang, Malaysia, 7–9 November 2019; pp. 99–104. [CrossRef]
37. Meng, Y.; Liu, X. Application of K-means Algorithm Based on Ant Clustering Algorithm in Macroscopic Planning of Highway Transportation Hub. In Proceedings of the 2007 First IEEE International Symposium on Information Technologies and Applications in Education, Kunming, China, 23–25 November 2007; pp. 483–488. [CrossRef]
38. Nath, R.P.D.; Lee, H.-J.; Chowdhury, N.K.; Chang, J.-W. Modified K-Means Clustering for Travel Time Prediction Based on Historical Traffic Data. In Proceedings of the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Cardiff, UK, 8–10 September 2010; pp. 511–521. [CrossRef]
39. Fernandez, S.; Ito, T. Driver Behavior Model Based on Ontology for Intelligent Transportation Systems. In *International Workshop on Learning Technology for Education in Cloud*; Springer: Cham, Switzerland, 2015; pp. 227–231. [CrossRef]
40. Fernandez, S.; Ito, T. Using SSN ontology for automatic traffic light settings on intelligent transportation systems. In Proceedings of the 2016 IEEE International Conference on Agents (ICA), Matsue, Japan, 28–30 September 2016; pp. 106–107.
41. Gregor, D.; Toral, S.; Ariza, T.; Barrero, F.; Gregor, R.; Rodas, J.; Arzamendia, M. A methodology for structured ontology construction applied to intelligent transportation systems. *Comput. Stand. Interfaces* **2016**, *47*, 108–119. [CrossRef]
42. Michalski, R.S.; Carbonell, J.G.; Mitchell, T.M. *Machine Learning: An Artificial Intelligence Approach*; Springer: New York, NY, USA, 2013; pp. 300–460.
43. Li, L.; Lv, Y.; Wang, F.-Y. Traffic signal timing via deep reinforcement learning. *IEEE/CAA J. Autom. Sin.* **2016**, *3*, 247–254. [CrossRef]
44. Arel, I.; Liu, C.; Urbanik, T.; Kohls, A. Reinforcement learning-based multi-agent system for network traffic signal control. *IET Intell. Transp. Syst.* **2010**, *4*, 128–135. [CrossRef]
45. Meng, X.; Bradley, J.; Yavuz, B.; Sparks, E.; Venkataraman, S.; Liu, D.; Freeman, J.; Tsai, D.B.; Amde, M.; Owen, S.; et al. Mllib: Machine learning in apache spark. *J. Mach. Learn. Res.* **2016**, *17*, 1–7.
46. Mian, R.; Ghanbari, H.; Zareian, S.; Shtern, M.; Litoiu, M. A Data Platform for the Highway Traffic Data. In Proceedings of the 2014 IEEE 8th International Symposium on the Maintenance and Evolution of Service-Oriented and Cloud-Based Systems, Raleigh, NC, USA, 29 September 2014; pp. 47–52. [CrossRef]

47. Shtern, M.; Mian, R.; Litoiu, M.; Zareian, S.; Abdelgawad, H.; Tizghadam, A. Towards a Multi-cluster Analytical Engine for Transportation Data. In Proceedings of the 2014 International Conference on Cloud and Autonomic Computing, Washington, DC, USA, 8–12 September 2014; pp. 249–257. [CrossRef]
48. Zareian, S.; Veleda, R.; Litoiu, M.; Shtern, M.; Ghanbari, H.; Garg, M. K-feed—A data-oriented approach to application performance management in cloud. In Proceedings of the 2015 IEEE 8th International Conference on Cloud Computing, New York, NY, USA, 27 June–2 July 2015; pp. 1045–1048.
49. Khazaei, H.; Zareian, S.; Veleda, R.; Litoiu, M. Sipresk: A Big Data Analytic Platform for Smart Transportation. In Proceedings of the Smart City 360° 2016. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Bratislava, Slovakia; Toronto, ON, Canada, 13–16 October 2015; Springer: Cham, Switzerland, 2015; Volume 166, pp. 419–430.
50. Kamel, I.R.; Abdelgawad, H.; Abdulhai, B. Transportation Big Data Simulation Platform for the Greater Toronto Area (GTA). In *Smart City 360°*; Springer: New York, NY, USA, 2016; pp. 443–454. [CrossRef]
51. Cheng, N.; Lyu, F.; Chen, J.; Xu, W.; Zhou, H.; Zhang, S.; Shen, X. Big Data Driven Vehicular Networks. *IEEE Netw.* **2018**, *32*, 160–167. [CrossRef]
52. Avdeeva, I.L.; Polyanin, A.V.; Golovina, T.A. Digitalization of industrial economic systems: Problems and consequences of modern technologies. *Econ. Control. Law* **2019**, *19*, 238–245. [CrossRef]
53. Hale, G.; Lopez, J.A. Monitoring banking system connectedness with big data. *J. Econ.* **2019**, *212*, 203–220. [CrossRef]
54. Hasan, M.; Popp, J.; Oláh, J. Current landscape and influence of big data on finance. *J. Big Data* **2020**, *7*, 21. [CrossRef]
55. Einav, L.; Levin, J. The Data Revolution and Economic Analysis. *Innov. Policy Econ.* **2014**, *14*, 1–24. [CrossRef]
56. Cheng, X.; Liu, S.; Sun, X.; Wang, Z.; Zhou, H.; Shao, Y.; Shen, H. Combating emerging financial risks in the big data era: A perspective review. *Fundam. Res.* **2021**, *1*, 595–606. [CrossRef]
57. Gates, C.; Matthews, P. Data is the new currency. In Proceedings of the 2014 New Security Paradigms Workshop; ACM: Victoria, Australia, 2014; pp. 105–116.
58. Shen, D.; Chen, S. Big data finance and financial markets. In Proceedings of the Big Data in Computational Social Science and Humanities; Chen, S.H., Ed.; Series: Computational Social Sciences. Springer: New York, NY, USA, 2018; pp. 235–248.
59. Sun, Y.; Shi, Y.; Zhang, Z. Finance Big Data: Management, Analysis, and Applications. *Int. J. Electron. Commer.* **2019**, *23*, 9–11. [CrossRef]
60. Sun, W.; Zhao, Y.; Sun, L. Big Data Analytics for Venture Capital Application: Towards Innovation Performance Improvement. *Int. J. Inf. Manag.* **2018**, *50*, 557–565. [CrossRef]
61. Tang, Y.; Xiong, J.J.; Luo, Y.; Zhang, Y.-C. How Do the Global Stock Markets Influence One Another? Evidence from Finance Big Data and Granger Causality Directed Network. *Int. J. Electron. Commer.* **2019**, *23*, 85–109. [CrossRef]
62. Yadegaridehkordi, E.; Nilashi, M.; Shuib, L.; Nasir, M.H.N.B.M.; Asadi, S.; Samad, S.; Awang, N.F. The impact of big data on firm performance in hotel industry. *Electron. Commer. Res. Appl.* **2019**, *40*, 100921. [CrossRef]
63. Wright, L.T.; Robin, R.; Stone, M.; Aravopoulou, E. Adoption of Big Data Technology for Innovation in B2B Marketing. *J. Bus. Bus. Mark.* **2019**, *26*, 281–293. [CrossRef]
64. Chai, S.; Shih, W. Why big data isn't enough. *MIT Sloan Manag. Rev.* **2017**, *58*, 57–61.
65. Loukis, E.; Kyriakou, N.; Pazalos, K.; Popa, S. Inter-organizational innovation and cloud computing. *Electron. Commer. Res.* **2016**, *17*, 379–401. [CrossRef]
66. Babu, M.M.; Rahman, M.; Alam, A.; Dey, B.L. Exploring big data-driven innovation in the manufacturing sector: Evidence from UK firms. *Ann. Oper. Res.* **2021**, 1–28. [CrossRef] [PubMed]
67. Pousttchi, K.; Hufenbach, Y. Engineering the Value Network of the Customer Interface and Marketing in the Data-Rich Retail Environment. *Int. J. Electron. Commer.* **2014**, *18*, 17–42. [CrossRef]
68. Grover, P.; Kar, A.K. Big Data Analytics: A Review on Theoretical Contributions and Tools Used in Literature. *Glob. J. Flex. Syst. Manag.* **2017**, *18*, 203–229. [CrossRef]
69. Lacity, M.; Willcocks, L. Business process outsourcing and dynamic innovation. *Strateg. Outsourcing Int. J.* **2014**, *7*, 66–92. [CrossRef]
70. Zhan, Y.; Tan, K.H.; Ji, G.; Chung, L.; Tseng, M. A big data framework for facilitating product innovation processes. *Bus. Process. Manag. J.* **2017**, *23*, 518–536. [CrossRef]
71. Wamba, S.F. Big data analytics and business process innovation. *Bus. Process. Manag. J.* **2017**, *23*, 470–476. [CrossRef]
72. Han, J.; Pei, J.; Tong, H. *Data Mining Concepts and Techniques*, 4th ed.; Elsevier: Amsterdam, The Netherlands, 2022.
73. Willis, G.; Tranos, E. Using 'Big Data' to understand the impacts of Uber on taxis in New York City. *Travel Behav. Soc.* **2021**, *22*, 94–107. [CrossRef]
74. Zhang, H.L.; Zhao, Y.; Pang, C.; He, J. Splitting Large Medical Data Sets Based on Normal Distribution in Cloud Environment. *IEEE Trans. Cloud Comput.* **2015**, *8*, 518–531. [CrossRef]
75. Lyko, K.; Nitzschke, M.; Ngomo, A.N. Big Data Acquisition. In *New Horizons for a Data-Driven Economy*; Cavanillas, J.M., Ed.; Springer Press: New York, NY, USA, 2016; pp. 39–61.
76. Coda, F.A.; Filho, D.J.S.; Junqueira, F.; Miyagi, P.E. Big Data Acquisition Architecture: An Industry 4.0 Approach. In *Technological Innovation for Life Improvement*; Camarinha-Matos, L., Farhadi, N., Lopes, F., Pereira, H., Eds.; Springer: New York, NY, USA, 2020; pp. 222–229.

77. Chang, W.; Boyd, D.; Levin, O. *NIST Big Data Interoperability Framework: Volume 6, Reference Architecture*; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2019. Available online: https://www.nist.gov/publications/nist-big-data-interoperability-framework-volume-6-reference-architecture?pub_id=918936 (accessed on 19 October 2022).
78. Jain, P.; Gyanchandani, M.; Khare, N. Big data privacy: A technological perspective and review. *J. Big Data* **2016**, *3*, 25. [CrossRef]
79. Rafiq, F.; Awan, M.J.; Yasin, A.; Nobanee, H.; Zain, A.M.; Bahaj, S.A. Privacy Prevention of Big Data Applications: A Systematic Literature Review. *SAGE Open* **2022**, *12*. [CrossRef]
80. Saheb, T.; Saheb, T. Understanding the development trends of big data technologies: An analysis of patents and the cited scholarly works. *J. Big Data* **2020**, *7*, 1–26. [CrossRef]
81. Vinoski, S. Advanced Message Queuing Protocol. *IEEE Internet Comput.* **2006**, *10*, 87–89. [CrossRef]
82. Saint-Andre, P.; Smith, K.; Tronçon, R. *XMPP: The Definitive Guide: Building Real-Time Applications with Jabber Technologies*, 1st ed.; O'Reilly Press: Sebastopol, CA, USA, 2009.
83. Moniruzzaman, A.B.; Hossain, S.A. NoSQL Database: New Era of Databases for Big Data Analytics-Classification, Characteristics and Comparison. *arXiv* **2013**, arXiv:1307.0191.
84. Hannah, S.; Deepa, A.J.; Chooralil, V.S.; BrillySangeetha, S.; Yuvaraj, N.; Raja, R.A.; Suresh, C.; Vignesh, R.; AbdullahR, Y.; Srihari, K.; et al. Blockchain-Based Deep Learning to Process IoT Data Acquisition in Cognitive Data. *BioMed Res. Int.* **2022**, *2022*, 5038851. [CrossRef]
85. Tittel, E. *Data Fabric for Dummies. Hitachi Vantara Special Edition*; John Wiley & Sons Press: Hoboken, NJ, USA, 2021.
86. Zhang, Y.; Yang, X.J.; Zhou, F. Disengagement Cause-and-Effect Relationships Extraction Using an NLP Pipeline. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 1–10. [CrossRef]
87. Cook, I. Apache Hive Language Manual. Apache Software Foundation. 2018. Available online: https://cwiki.apache.org/confluence/display/Hive/LanguageManual (accessed on 9 December 2022).
88. IBM Cloud Education. Hadoop vs. Spark: What's the Difference? IBM Company Press. 2021. Available online: https://www.ibm.com/cloud/blog/hadoop-vs-spark (accessed on 30 November 2022).
89. Chen, S.; Haziza, D.; Mashreghi, Z. A Comparison of Existing Bootstrap Algorithms for Multi-Stage Sampling Designs. *Stats* **2022**, *5*, 521–537. [CrossRef]
90. Fu, Y.; Xiao, N.; Chen, T.; Wang, J. Fog-to-MultiCloud Cooperative Ehealth Data Management with Application-Aware Secure Deduplication. *IEEE Trans. Dependable Secur. Comput.* **2022**, *19*, 3136–3148. [CrossRef]
91. Zheng, Z.; Zhu, J.; Lyu, M.R. Service-Generated Big Data and Big Data-as-a-Service: An Overview. In Proceedings of the 2013 IEEE International Congress on Big Data, Silicon Valley, CA, USA, 6–9 October 2013; pp. 403–410. [CrossRef]
92. Lu, J.; Shen, J.; Vijayakumar, P.; Gupta, B.B. Blockchain-Based Secure Data Storage Protocol for Sensors in the Industrial Internet of Things. *IEEE Trans. Ind. Inform.* **2021**, *18*, 5422–5431. [CrossRef]
93. Tallis, B. NVM Express, Inc. Documentation. NVMe 1.4 Specification Published: Further Optimizing Performance and Reli-Ability. Available online: https://www.anandtech.com/show/14543/nvme-14-specification-published (accessed on 10 December 2022).
94. Huang, Q.; Zhang, H.L.; Chen, G.; Tong, J.; Lee, S. An Optimized High Dimension Data Reduction Method based on Covariance. In Proceedings of the Web Intelligence Workshops (WI), Niagara Falls, ON, Canada, 17–20 November 2022; pp. 1–6.
95. Niemann, K.; Wolpers, M. A new collaborative filtering approach for increasing the aggregate diversity of recommender systems. In Proceedings of the 19th ACM SIGKDD on Knowledge Discovery and Data mining (KDD), Chicago, IL, USA, 11–14 August 2013; pp. 955–963. [CrossRef]
96. Winlaw, M.; Hynes, M.B.; Caterini, A.; De Sterck, H. Algorithmic Acceleration of Parallel ALS for Collaborative Filtering: Speeding up Distributed Big Data Recommendation in Spark. In Proceedings of the IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS), Melbourne, Australia, 14–17 December 2015; pp. 682–691. [CrossRef]
97. Kriegel, H.; Kröger, P.; Sander, J.; Zimek, A. Density-based clustering. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2011**, *1*, 231–240. [CrossRef]
98. Akinduko, A.A.; Mirkes, E.; Gorban, A. SOM: Stochastic initialization versus principal components. *Inf. Sci.* **2016**, *364–365*, 213–221. [CrossRef]
99. Sangeetha, K.; Prabha, D. RETRACTED ARTICLE: Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *12*, 4117–4126. [CrossRef]
100. Rani, K.S.; Kumari, M.; Singh, V.B.; Sharma, M. Deep Learning with Big Data: An Emerging Trend. In Proceedings of the 19th International Conference on Computational Science and Its Applications, Saint Petersburg, Russia, 1–4 July 2019; pp. 93–101.
101. Moran, A.; Gadepally, V.; Hubbell, M.; Kepner, J. Improving Big Data visual analytics with interactive virtual reality. In Proceedings of the IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, 15–17 September 2015; pp. 1–6.
102. Im, D.; Lee, S.; Lee, H.; Yoon, B.; So, F.; Jeong, J. A Data-Centric Approach to Design and Analysis of a Surface-Inspection System Based on Deep Learning in the Plastic Injection Molding Industry. *Processes* **2021**, *9*, 1895. [CrossRef]
103. Salomaa, A. *Public-Key Cryptography*, 2nd ed.; Springer: Berlin, Germany, 1996; p. 275.
104. Xu, K.; Yue, H.; Guo, L.; Guo, Y.; Fang, Y. Privacy-Preserving Machine Learning Algorithms for Big Data Systems. In Proceedings of the IEEE 35th International Conference on Distributed Computing Systems, Columbus, OH, USA, 29 June–2 July 2015; pp. 318–327.

105. Dwork, C.; Roth, A. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.* **2014**, *9*, 211–407. [CrossRef]
106. Bhathal, G.S.; Singh, A. Big Data: Hadoop framework vulnerabilities, security issues and attacks. *Array* **2019**, *1–2*. [CrossRef]
107. Jagatheesaperumal, S.K.; Rahouti, M.; Ahmad, K.; Al-Fuqaha, A.; Guizani, M. The Duo of Artificial Intelligence and Big Data for Industry 4.0: Applications, Techniques, Challenges, and Future Research Directions. *IEEE Internet Things J.* **2021**, *9*, 12861–12885. [CrossRef]
108. Wang, X.; Yang, L.T.; Liu, H.; Deen, M.J. A Big Data-as-a-Service Framework: State-of-the-Art and Perspectives. *IEEE Trans. Big Data* **2017**, *4*, 325–340. [CrossRef]
109. Oseledets, I. Tensor-Train Decomposition. *SIAM J. Sci. Comput.* **2011**, *33*, 2295–2317. [CrossRef]
110. Dionisio, J.; Gilbert, R. 3D virtual worlds and the Metaverse: Current status and future possibilities. *ACM Comput. Surv.* **2013**, *45*, 1–38. [CrossRef]
111. Suzuki, S.-N.; Kanematsu, H.; Barry, D.M.; Ogawa, N.; Yajima, K.; Nakahira, K.T.; Shirai, T.; Kawaguchi, M.; Kobayashi, T.; Yoshitake, M. Virtual Experiments in Metaverse and their Applications to Collaborative Projects: The framework and its significance. *Procedia Comput. Sci.* **2020**, *176*, 2125–2132. [CrossRef]
112. Gogolin, G.; Gogolin, E.; Kam, H.-J. Virtual worlds and social media: Security and privacy concerns, implications, and practices. *Int. J. Artif. Life Res.* **2014**, *4*, 30–42. [CrossRef]
113. Mozumder, M.A.I.; Sheeraz, M.M.; Athar, A.; Aich, S.; Kim, H.-C. Overview: Technology Roadmap of the Future Trend of Metaverse based on IoT, Blockchain, AI Technique, and Medical Domain Metaverse Activity. In Proceedings of the 24th Interna-tional Conference on Advanced Communication Technology, Pyeongchang, Korea, 13–16 February 2022; pp. 256–261.
114. de Gérase, N. *Nicomachi Geraseni Pythagorei Introductionis Arithmeticae Libri II*; Bibliotheca Scriptorum Graecorum et Romanorum Teubneriana; Aedibvs B.G. Teubneri: Stuttgart, Germany, 1866; pp. 1–198. Available online: https://openlibrary.org/works/OL3947510W/Nicomachi_Geraseni_Pythagorei_introductionis_arithmeticae_libri_II (accessed on 30 November 2022).
115. Rivest, R.L.; Shamir, A.; Adleman, L. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM* **1978**, *21*, 120–126. [CrossRef]
116. Kolotouros, N.; Pavlakos, G.; Black, M.; Daniilidis, K. Learning to Reconstruct 3D Human Pose and Shape via Model-Fitting in the Loop. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019. [CrossRef]
117. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 172–186. [CrossRef] [PubMed]
118. Lin, S.; Yang, L.; Saleemi, I.; Sengupta, S. Robust high-resolution video matting with temporal guidance. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4 – 8 January 2022; pp. 238–247.
119. Bhatia, S.; Sigal, L.; Isard, M.; Black, M. 3D Human Limb Detection using Space Carving and Multi-View Eigen Models. In Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop, Washington, DC, USA, 27 June–2 July 2004; p. 17.
120. Agarwal, A.; Triggs, B. Recovering 3D Human Pose from Monocular Images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 44–58. [CrossRef]
121. Martinez, J.; Hossain, R.; Romero, J.; Little, J.J. A Simple Yet Effective Baseline for 3D Human Pose Estimation. In Proceedings of the IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, 22–29 October 2017; pp. 2659–2668.
122. Mehta, D.; Sotnychenko, O.; Mueller, F.; Xu, W.; Elgharib, M.; Fua, P.; Seidel, H.P.; Rhodin, H.; Pons-Moll, G.; Theobalt, C. XNect: Real-time multi-person 3D motion capture with a single RGB camera. *ACM Trans. Graph.* **2020**, *39*, 82:1–82:17. [CrossRef]
123. Federica, B.; Kanazawa, A.; Lassner, C.; Gehler, P.; Romero, J.; Black, M.J. Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 561–578.
124. Kanazawa, A.; Black, M.J.; Jacobs, D.W.; Malik, J. End-to-end Recovery of Human Shape and Pose. In Proceedings of the Computer Vision and Pattern Regognition (CVPR), San Juan, PR, USA, 17–19 June 2018; pp. 1–10.
125. Pavlakos, G.; Zhu, L.; Zhou, X.; Daniilidis, K. Learning to Estimate 3D Human Pose and Shape from a Single Color Image. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 459–468.
126. Pavlakos, G.; Choutas, V.; Ghorbani, N.; Bolkart, T.; Osman, A.A.; Tzionas, D.; Black, M.J. Expressive body capture: 3d hands, face, and body from a single image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2019; pp. 10975–10985.
127. Choutas, V.; Pavlakos, G.; Bolkart, T.; Tzionas, D.; Black, M.J. Monocular Expressive Body Regression Through Body-Driven Attention. In Proceedings of the European Conference on Computer Vision; Springer: Cham, Switzerland, 2020; pp. 20–40. [CrossRef]
128. Zhang, Y.; Li, Z.; An, L.; Li, M.; Yu, T.; Liu, Y. Lightweight multi-person total motion capture using sparse multi-view cameras. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 5560–5569.
129. Zheng, Z.; Yu, T.; Liu, Y.; Dai, Q. Pamir: Parametric model-conditioned implicit representation for image-based human reconstruction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3170–3184. [CrossRef]
130. Huang, Z.; Xu, Y.; Lassner, C.; Li, H.; Tung, T. Arch: Animatable reconstruction of clothed humans. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, DC, USA, 13–19 June 2020; pp. 3093–3102.

131. Ma, Q.; Yang, J.; Tang, S.; Black, M.J. The power of points for modeling humans in clothing. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10974–10984.
132. Peng, S.; Zhang, Y.; Xu, Y.; Wang, Q.; Shuai, Q.; Bao, H.; Zhou, X. Neural Body: Implicit Neural Representations with Structured Latent Codes for Novel View Synthesis of Dynamic Humans. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 9050–9059. [CrossRef]
133. Mir, A.; Alldieck, T.; Pons-Moll, G. Learning to transfer texture from clothing images to 3d humans. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, DC, USA, 13–19 June 2020; pp. 7023–7034.
134. Yang, F.; Li, R.; Georgakis, G.; Karanam, S.; Chen, T.; Ling, H.; Wu, Z. Robust multi-modal 3d patient body modeling. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention; Springer: New York, NY, USA, 2020; pp. 86–95.
135. Lee, H.-J.; Chen, Z. Determination of 3D human body postures from a single view. *Comput. Vis. Graph. Image Process.* **1985**, *30*, 148–168. [CrossRef]
136. Nevatia, R.; Binford, T.O. Description and recognition of curved objects. *Artif. Intell.* **1977**, *8*, 77–98. [CrossRef]
137. Ju, S.X.; Black, M.J.; Yacoob, Y. Cardboard people: A parameterized model of articulated image motion. In Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, Killington, VT, USA, 14–16 October 1996; pp. 38–44.
138. Wang, M.; Qiu, F.; Liu, W.; Qian, C.; Zhou, X.; Ma, L. Monocular Human Pose and Shape Reconstruction using Part Differen-tiable Rendering. *Comput. Graph. Forum* **2020**, *39*, 351–362. [CrossRef]
139. Robinette, K.M.; Blackwell, S.; Daanen, H.; Boehmer, M.; Fleming, S. Civilian American and European Surface Anthropometry Resource (CAESAR), Final Report. Volume 1. Summary. 2002. Available online: https://www.humanics-es.com/CAESARvol1.pdf (accessed on 30 November 2022).
140. Anguelov, D.; Srinivasan, P.; Koller, D.; Thrun, S.; Rodgers, J.; Davis, J. SCAPE: Shape completion and animation of people. *ACM Trans. Graph.* **2005**, *24*, 408–416. [CrossRef]
141. Loper, M.; Mahmood, N.; Romero, J.; Pons-Moll, G.; Black, M.J. SMPL: A Skinned Multi-Person Linear Model. *ACM Trans. Graphics* **2015**, *34*, 248. [CrossRef]
142. Romero, J.; Tzionas, D.; Black, M.J. Embodied hands: Modeling and capturing hands and bodies together. *arXiv* **2022**, arXiv:2201.02610. [CrossRef]
143. Blanz, V.; Vetter, T. A morphable model for the synthesis of 3D faces. In *Siggraph 1999, Computer Graphics Proceedings*; Rockwood, A., Ed.; Addison Wesley Longman: Los Angeles, CA, USA, 1999; pp. 187–194. [CrossRef]
144. Osman, A.A.A.; Bolkart, T.; Black, M.J. STAR: A Sparse Trained Articulated Human Body Regressor. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; pp. 598–613.
145. Saito, S.; Huang, Z.; Natsume, R.; Morishima, S.; Li, H.; Kanazawa, A. PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 2304–2314.
146. Saito, S.; Simon, T.; Saragih, J.; Joo, H. PIFuHD: Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization. In Proceedings of the 2020 IEEE/CVF Conference on CVPR, Seattle, DC, USA, 13–19 June 2020; pp. 81–90. [CrossRef]
147. Balan, A.O.; Sigal, L.; Black, M.J.; Davis, J.E.; Haussecker, H.W. Detailed Human Shape and Pose from Images. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8. [CrossRef]
148. Guan, P.; Weiss, A.; Balan, A.O.; Black, M.J. Estimating human shape and pose from a single image. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 1381–1388.
149. Pishchulin, L.; Insafutdinov, E.; Tang, S.; Andres, B.; Andriluka, M.; Gehler, P.V.; Schiele, B. DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 4929–4937.
150. Zanfir, A.; Marinoiu, E.; Sminchisescu, C. Monocular 3D Pose and Shape Estimation of Multiple People in Natural Scenes: The Importance of Multiple Scene Constraints. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2148–2157. [CrossRef]
151. Christoph, L.; Romero, J.; Kiefel, M.; Bogo, F.; Black, M.J.; Gehler, P.V. Unite the people: Closing the loop between 3d and 2d human representations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6050–6059.
152. Xiang, D.; Joo, H.; Sheikh, Y. Monocular total capture: Posing face, body, and hands in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
153. Joo, H.; Neverova, N.; Vedaldi, A. Exemplar Fine-Tuning for 3D Human Model Fitting Towards In-the-Wild 3D Human Pose Estimation. In Proceedings of the 2021 International Conference on 3D Vision (3DV), Virtual, 1–3 December 2021; pp. 42–52. [CrossRef]
154. Song, J.; Chen, X.; Hilliges, O. Human body model fitting by learned gradient descent. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Cham, Switzerland, 2020; pp. 744–760.
155. Iqbal, U.; Xie, K.; Guo, Y.; Kautz, J.; Molchanov, P. KAMA: 3D Keypoint Aware Body Mesh Articulation. In Proceedings of the 2021 International Conference on 3D Vision (3DV), Virtual, 1–3 December 2021; pp. 689–699. [CrossRef]

156. Li, J.; Xu, C.; Chen, Z.; Bian, S.; Yang, L.; Lu, C. HybrIK: A Hybrid Analytical-Neural Inverse Kinematics Solution for 3D Human Pose and Shape Estimation. In Proceedings of the IEEE/CVF CVPR, Virtual, 20–25 June 2021; pp. 3382–3392. [CrossRef]
157. Jiang, W.; Kolotouros, N.; Pavlakos, G.; Zhou, X.; Daniilidis, K. Coherent reconstruction of multiple humans from a single image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, DC, USA, 13–19 June 2020; pp. 5579–5588.
158. Kolotouros, N.; Pavlakos, G.; Daniilidis, K. Convolutional mesh regression for single-image human shape reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4501–4510.
159. Kocabas, M.; Athanasiou, N.; Black, M.J. VIBE: Video Inference for Human Body Pose and Shape Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, DC, USA, 13–19 June 2020; pp. 5252–5262.
160. Mahmood, N.; Ghorbani, N.; Troje, N.F.; Pons-Moll, G.; Black, M.J. AMASS: Archive of Motion Capture as Surface Shapes. In Proceedings of the International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 5442–5451.
161. Feng, Y.; Choutas, V.; Bolkart, T.; Tzionas, D.; Black, M.J. Collaborative regression of expressive bodies using moderation. In Proceedings of the 2021 International Conference on 3D Vision (3DV), London, UK, 1–3 December 2021; pp. 792–804.
162. Zanfir, A.; Bazavan, E.G.; Zanfir, M.; Freeman, W.T.; Sukthankar, R.; Sminchisescu, C. Neural descent for visual 3d human pose and shape. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14484–14493.
163. Zhou, Y.; Habermann, M.; Habibie, I.; Tewari, A.; Theobalt, C.; Xu, F. Monocular Real-Time Full Body Capture with Inter-Part Correlations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 4811–4822.
164. Moon, G.; Lee, K.M. Pose2pose: 3d positional pose-guided 3d rotational pose prediction for expressive 3d human pose and mesh estimation. *arXiv* **2020**, arXiv:2011.11534.
165. Wang, Y.; Su, Z.; Zhang, N.; Xing, R.; Liu, D.; Luan, T.H.; Shen, X. A Survey on Metaverse: Fundamentals, Security, and Privacy. In *IEEE Communications Surveys & Tutorials 2022*; IEEE Press: New York, NY, USA, 2022; pp. 1–32.
166. Huang, J.; Zhang, H.L.; Lu, H.; Xin, Y.; Li, S. A Novel Position-based VR Online Shopping Recommendation System based on Optimized Collaborative Filtering Algorithm. In Proceedings of the Web Intelligence Workshops (WI), Melbourne, VIC, Canada, 14–17 December 2022; pp. 1–7.
167. Fuller, A.; Fan, Z.; Day, C.; Barlow, C. Digital Twin: Enabling Technologies, Challenges and Open Research. *IEEE Access* **2020**, *8*, 108952–108971. [CrossRef]
168. Carver, J.; Staron, M.; Capilla, R.; Muccini, H.; Hochstein, L. Digital Twins. *IEEE Softw.* **2022**, *39*, 97–99. [CrossRef]
169. Zhang, H.L.; Liu, J.; Dowens, M.G. Complex brain activity analysis and recognition based on multiagent methods. *Concurr. Comput. Pr. Exp.* **2020**, *34*, e5855. [CrossRef]
170. Zhang, H.; Zhao, Q.; Lee, S.; Dowens, M.G. EEG-Based Driver Drowsiness Detection Using the Dynamic Time Dependency Method. In Proceedings of the Brain Informatics, Haikou, China, 13–15 December 2019; pp. 39–47. [CrossRef]
171. Zhang, H.L.; Lee, S.; Li, X.; He, J. EEG Self-Adjusting Data Analysis Based on Optimized Sampling for Robot Control. *Electronics* **2020**, *9*, 925. [CrossRef]
172. Xu, H.; Li, Z.; Li, Z.; Zhang, X.; Sun, Y.; Zhang, L. Metaverse Native Communication: A Blockchain and Spectrum Prospective. In Proceedings of the IEEE International Conference on Communications Workshops, Seoul, Korea, 16–20 May 2022; pp. 7–12.
173. Narayanan, A.; Bonneau, J.; Felten, E.; Miller, A.; Goldfeder, S. *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction*; Princeton University Press: Princeton, NJ, USA, 2016.
174. Talamo, M.; Arcieri, F.; Dimitri, A.; Schunck, H.C. A blockchain based PKI validation system based on rare events manage-ment. *Future Internet* **2020**, *12*, 40. [CrossRef]
175. Badruddoja, S.; Dantu, R.; He, Y.; Thompson, M.; Salau, A.; Upadhyay, K. *Trusted AI with Blockchain to Empower Metaverse. International Conference on Blockchain Computing and Applications (BCCA)*; IEEE Press: New York, NY, USA, 2022; pp. 237–245.
176. Mühle, A.; Grüner, A.; Gayvoronskaya, T.; Meinel, C. A survey on essential components of a self-sovereign identity. *Comput. Sci. Rev.* **2018**, *30*, 80–86. [CrossRef]
177. Heimes, A.; Zenkert, J.; Fathi, M. Current State and Latest Trends in Blockchain Technology and its Usage and the Effects on Business Use Cases. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), Melbourne, Australia, 17–20 October 2021; pp. 3311–3316. [CrossRef]
178. Chang, L.; Zhang, Z.; Li, P.; Xi, S.; Guo, W.; Shen, Y.; Xiong, Z.; Kang, J.; Niyato, D.; Qiao, X.; et al. 6G-Enabled Edge AI for Metaverse: Challenges, Methods, and Future Research Directions. *J. Commun. Inf. Netw.* **2022**, *7*, 107–121. [CrossRef]
179. Prayitno; Shyu, C.-R.; Putra, K.T.; Chen, H.-C.; Tsai, Y.-Y.; Hossain, K.S.M.T.; Jiang, W.; Shae, Z.-Y. A Systematic Review of Federated Learning in the Healthcare Area: From the Perspective of Data Properties and Applications. *Appl. Sci.* **2021**, *11*, 11191. [CrossRef]
180. Bhugaonkar, K.; Bhugaonkar, R.; Masne, N. The Trend of Metaverse and Augmented & Virtual Reality Extending to the Healthcare System. *Cureus* **2022**, *14*, 29071. [CrossRef]
181. Sun, H.; Rabbani, M.R.; Sial, M.S.; Yu, S.; Filipe, J.A.; Cherian, J. Identifying Big Data's Opportunities, Challenges, and Implications in Finance. *Mathematics* **2020**, *8*, 1738. [CrossRef]

182. Musamih, A.; Dirir, A.; Yaqoob, I.; Salah, K.; Jayaraman, R.; Puthal, D. NFTs in Smart Cities: Vision, Applications, and Chal-lenges. *IEEE Consum. Electron. Mag.* **2022**, 1–14. [CrossRef]
183. Edwards, C. Are NFTs Key to Accessing the Metaverse? *Eng. Technol.* **2022**, *17*, 1–8. [CrossRef]
184. Shabihi, N.; Kim, M.S. Big Data Analytics in Education: A Data-Driven Literature Review. In Proceedings of the International Conference on Advanced Learning Technologies (ICALT), Tartu, Estonia, 12–15 July 2021; pp. 154–156.
185. Qureshi, H.; Sagar, A.K.; Astya, R.; Shrivastava, G. Big Data Analytics for Smart Education. In Proceedings of the IEEE 6th Inter-national Conference on Computing, Communication and Automation (ICCCA), Arad, Romania, 17–19 December 2021; pp. 650–658.
186. Gim, G.; Bae, H.; Kang, S. Metaverse Learning: The Relationship among Quality of VR-Based Education, Self-Determination, and Learner Satisfaction. In Proceedings of the IEEE/ACIS 7th International Conference on Big Data, Cloud Computing, and Data Science, Danang, Vietnam, 4–6 October 2022; pp. 279–284. [CrossRef]
187. Agrawal, D.; Budak, C.; El Abbadi, A.; Georgiou, T.; Yan, X. Big Data in Online Social Networks: User Interaction Analysis to Model User Behavior in Social Networks. In Proceedings of the Databases in Networked Information Systems, Aizu-Wakamatsu, Japan, 24–26 March 2014; Madaan, A., Kikuchi, S., Bhalla, S., Eds.; Lecture Notes in Computer Science. pp. 1–16. [CrossRef]
188. Britto, L.F.S.; Pacifico, L.D.S. Evaluating Video Game Acceptance in Game Reviews using Sentiment Analysis Techniques. In Proceedings of the SBGames, Virtual, 7–10 November 2020; pp. 399–402.
189. Mirza-Babaei, P.; Robinson, R.; Mandryk, R.; Pirker, J.; Kang, C.; Fletcher, A. Games and the Metaverse. In Proceedings of the Annual Symposium on Computer-Human Interaction in Play, Bremen, Germany, 2–5 November 2022; pp. 318–319.
190. Cheng, R.; Wu, N.; Varvello, M.; Chen, S.; Han, B. Are we ready for Metaverse: A measurement study of social virtual reality platforms. In Proceedings of the 22nd ACM Internet Measurement Conference, Nice, France, 25–27 October 2022; pp. 504–518.
191. Templier, M.; Paré, G. A framework for guiding and evaluating literature reviews. *Commun. Assoc. Inf. Syst.* **2015**, *37*, 112–137. [CrossRef]
192. Snyder, H. Literature review as a research methodology: An overview and guidelines. *J. Bus. Res.* **2019**, *104*, 333–339. [CrossRef]
193. Chen, C. How to Use CiteSpace (6.1.R5) Advanced. CiteSpace. 2020. Available online: https://citespace.podia.com/ebook-how-to-use-citespace (accessed on 30 November 2022).
194. Hare, P. Literature Reviews: Google Scholar. Bank Street College of Education. 2020. Available online: https://www.bankstreet.edu/library/blog/google-scholar/ (accessed on 30 November 2022).
195. Suzanne, F. Semantic Scholar. *J. Med. Libr. Assoc.* **2018**, *106*, 145–147.
196. Hu, X.; Liu, Y.; Zhang, H.L.; Wang, W.; Li, Y.; Meng, C.; Fu, Z. Noninvasive Human-Computer Interface Methods and Applications for Robotic Control: Past, Current, and Future. *Comput. Intell. Neurosci.* **2022**, *2022*, 1635672.
197. Hobson, D. How Banks Can Make Money in the Metaverse. *Future of Finance*, 17 June 2022; 1–4.
198. Falchuk, B.; Loeb, S.; Neff, R. The Social Metaverse: Battle for Privacy. *IEEE Technol. Soc. Mag.* **2018**, *37*, 52–61. [CrossRef]
199. Abiri, R.; Borhani, S.; Sellers, E.W.; Jiang, Y.; Zhao, X. A comprehensive review of EEG-based brain-computer interface par-adigms. *J. Neural Eng.* **2019**, *16*, 1–21. [CrossRef] [PubMed]
200. Oudeyer, P.Y.; Gottlieb, J.; Lopes, M. Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. *Prog. Brain Res.* **2016**, *229*, 257–284. [PubMed]
201. Fahimi, F.; Dosen, S.; Ang, K.K.; Mrachacz-Kersting, N.; Guan, C. Generative Adversarial Networks-Based Data Augmentation for Brain–Computer Interface. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4039–4051. [CrossRef] [PubMed]
202. Chen, X.; Huang, X.; Wang, Y.; Gao, X. Combination of Augmented Reality Based Brain- Computer Interface and Computer Vision for High-Level Control of a Robotic Arm. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 3140–3147. [CrossRef]