UNIVERSITI TEKNOLOGI MARA

ICT TRENDS THAT MATTER

Shamsatun Nahar Ahmad
Noor Azrin Zainuddin
Basri Badyalina
Nur Azlina Mat Noor
Muhammad Zulqarnain Hakim Abd. Jalal
Faten Elina Kamaruddin
Nurul Huda Md Yatim

FAKULTI SAINS KOMPUTER DAN MATEMATIK
UNIVERSITI TEKNOLOGI MARA
CAWANGAN JOHOR

Terbitan Edisi 2025

e ISBN: 978-629-7647-05-0

Diterbitkan oleh:
Universiti Teknologi MARA Cawangan Johor
Jalan Universiti Off KM 12 Jalan Muar ,

85000 Segamat, Johor .
Tel: 07-9352000
Fax: 07-9352716
https://johor .uitm.edu.my

EDITORIAL BOARD

TABLE OF CONTENTS

Preface	vi
Synopsis	vii
Acknowledgement	viii
Understanding and Utilizing Social Media Analytics Tools	1
Towards a Smart and Data-Driven Campus: Digital Ecosystem Development and The Uitm Johor RSP16 Experience	4
Telegram: 9 Reasons Why We Should Use It?	8
Tiktok Goes Global	11
Teaching & Learning: From Rubrics to Comprehensive Reports	15
Fun & Free E-Learning Apps For Kids: Making Learning an Adventure!	17
Swot Analysis of Chatgpt and Siri: Understanding Their Role and Impact as Popular Ai Tools	23
Improving Conceptual Understanding of Topics in Calculus via V-Ccmpedia	30
Ai Tools That “Wow” Your Students for Better Engagement in the Classroom	35
Computer Tips and Tricks: How to Make Your Pc Run Faster	41
Easymath2u: Learning Mathematics Beyond the Classroom	47
The Role of Artificial Intelligence (Ai) in Cybersecurity: Threats and Defenses	52
Big Data Harmonization for Enhanced Efficiency in Real-World Applications	59
Index	68

PREFACE

Praise be to Allah SWT, with His will, this eBook, ICT Trends that Matter, has been successfully compiled to capture some of the most relevant and transformative discussions in the world of Information and Communication Technology (ICT).

The work is a compilation of various views of the different practitioners, scholars, and professionals who have contributed their ideas and thoughts regarding the emerging technologies and their influence. The chapters provide just a few examples of how cybersecurity, big data harmonisation, artificial intelligence, novel learning tools, and social media analytics demonstrate the extent to which ICT has permeated our everyday worlds, our classrooms, workplaces, and communities.

ICT Trends that Matter offers readers a comprehensive exploration of 14 contemporary ICT themes that are shaping education, industry, and society. The eBook covers a wide spectrum of topics such as Big Data & AI, Digital Learning & Tools, Practical ICT Applications, Social Media & Communication and Smart Campus Initiatives highlighting UiTM Johor's experience in developing a data-driven digital ecosystem.

This eBook is informative and inspirational, with contributions that combine theory, research, and practical work. It makes the readers consider the existing ICT issues and opportunities and provides practical knowledge on personal, educational, and professional development. I would like to say that I am very grateful as the chief editor to all thecontributors whose commitment, professionalism, and innovativeness have added value to the contents of this eBook. I believe ICT Trends that Matter will be useful to academicians and students, as well as any industry professional, policymaker and those who are keen to learn more about the dynamic ICT environment.

Whether you are an academic, student, or industry professional, ICT Trends that Matter provides valuable insights into the technologies that are redefining our world today. May this work inspire further dialogue, innovation, and collaboration toward building a smarter and more sustainable digital future.

Dr. Shamsatun Nahar Ahmad
Chief Editor
Brain Hub: ICT Trends that Matter

SYNOPSIS

ICT Trends that Matter is a compilation of 14 thought-provoking chapters, which discuss the most significant trends in Information and Communication Technology (ICT) and their implications on education, industry, and society.

The elements cut across essential areas of the digital world. Discussions about the harmonisation of big data and artificial intelligence to fight cybersecurity and comparative studies concerning popular AI tools will be available to the readers.The eBook also highlights innovative approaches to teaching and learning, such as Easymath2U and V-CCMPedia, to improve conceptual learning in calculus, and AI-assisted tools to improve student engagement.

The useful experience is presented with the help of the following topics: computer tips and tricks, free e-learning applications used by children, and the successful utilisation of social media analytics tools. The role of contemporary communication mediums such as Telegram and the global presence of TikTok are also discussed in the chapters, as well as reflections on institutional work towards data-driven digital ecosystems, such as the UiTM Johor RSP16 experience.

This eBook contains the work of numerous scholars and researchers and offers both theoretical insights and practical solutions, which is why it can be of interest to academics, students, practitioners in the industry, or policymakers. ICT Trends that Matter is not merely an anthology of articles but rather is a convenient way to learn about the latest trends in ICT and predict what to expect and what to take advantage of in the digital age.

ACKNOWLEDGEMENT

The Editorial Board of ICT Trends that Matter would like to thank everyone whose assistance and commitment enabled us to make this publication possible.

We would like to thank the Department of Linkage Industry and Alumni, UiTM Johor, Segamat Campus, for enabling the acquisition of eISBN and subsequent guidance throughout the publication process.

A special mention of gratitude belongs to all contributors, whose skills, knowledge and dedication have been instrumental in the content of this eBook. Every chapter is an embodiment of how well, creatively, and committed our writers were to delivering substantial discussions on the current issues in ICT.

We also recognise the unwearying efforts on the part of the Editorial Board, which have been tireless from the very beginning of the conception to the final production of this eBook, which makes it and guarantees its success.

We are most thankful to all who have assisted this undertaking either directly or indirectly. May Allah SWT bless this endeavour and enable it to do good for the readers and the community at large.

BIG DATA HARMONISATION FOR ENHANCED EFFICIENCY IN REAL-WORLD APPLICATIONS

SITI HAJAR BAHARIN, SITI NUR SHAHIRA DAHARI,
NURFAIZAH KAMARUDDIN, RABIATUL ADAWIYAH KAMARULZAMAN

Introduction

Over four hundred million terabytes of data are generated daily. In the modern era, managing information from various sources poses a challenge because of its scale and speed (Marr, 2016; IDC, 2021; Muhammad, 2025). In 2024, for instance, there were roughly 5.9 million Google searches, 16,000 TikTok videos uploaded, and over 3.4 million YouTube views per minute, according to Eaves (2025). This sets new boundaries for businesses given the amount of semi-structured, unstructured, or structured data gathered from social platforms, machines, or transaction-based systems using traditional tools. Paying via apps, airline check-ins, and even medical record-keeping enable Internet of Things (IoT) devices such as smartwatches to gather valuable stream data online. The biggest blocker in such cases is the heterogeneity that big data presents because it includes numerous frameworks where standards, along with their accompanying qualities, are disparate. When experts are provided with vast amounts of fragmented information due to gaps created by silos or incompatible formats within data separated into chunks without any relation to each other's reference system, they become "blind and deaf".

Without harmonisation, huge unexplored datasets lose their potential insights and analysis due to fragmentation alongside inconsistency problems which arise when attempting to combine differing standards or extracts of sets sharing no relations with one another but claiming to hold value without comparability. Therefore, this paper focuses on exploring core capabilities of big data by unlocking extra layers of everyday applications through proper implementation which gradually unfolds when advanced techniques are applied for turning raw datasets ready to be processed into priceless treasures exhibiting knowledge (SAP, 2023). Intelligent and tailored functions of everyday domains such as fitness coaching, smart transportation, entertainment streaming, and online shopping recommendations are examples of where synergised big data is applied. Big data harmonization issues include differing data formats, integration difficulties, inconsistent data quality, and concerns over privacy. This paper outlines strategies detail best practices that address these challenges and still reach the conclusion that big data harmonisation will persist as a cornerstone process in the artificial intelligence (AI) and big data era, enabling more effective data-driven applications.

Big Data: Definition and Sources

Big data indicates data sets which are extensive, intricate, and diverse. They pose significant challenges to traditional tools used for their storage, retrieval, and processing. Expanding on the understanding of big data resulted in increasing the number of its "Vs", describing its composed features from 42 Vs (Shafer, 2017) to 51

Vs (Khan et al., 2019), with its additional ones being: veracity, variability, visualisation, value, and viability, accentuating its importance concerning its impact around a given organisational structure. Understanding big data requires identifying its primary sources which include:

- *Participation of Users and Social Platforms:* User participation through social applications like Facebook and Instagram creates vast repositories of unstructured information as a result of user engagement, thus creating high-volume bulk data while also accelerating (velocity) trends that underpin behaviours.
- *Transactional Data:* This encompasses information contained within traditional/online transactions such as banking systems or e-commerce websites which hold paramount importance towards business analytics, revealing economic trends alongside purchasing patterns.
- *IoT and Machine-Generated Data:* Classification of machine-generated data has become rampant owing to internet-inclined devices, culminating to collection of incessant streams of useful information which greatly contributes towards big data across sectors.
- *Media, Communications, along with Metadata Telecommunication companies* aided through large-scale user-financed telecasting services coupled with telecom companies gathering metadata, have greatly contributed towards the making of (Internet) big /large data through streaming file outputs creating meta reports retrievable by analysing datasets brimming in size. (This bulleted point is not using the same structure as the other 4 bulleted points where for the other 4 the structure is a phrase + colon (example: Open Data and Public Records:)
- *Open Data and Public Records*: Government and scientific organisations provide public datasets that can be combined with commercial data for insights, such as predicting traffic patterns using weather data.
-

Combining privately held information with publicly available datasets released by government agencies or scientific bodies offers valuable analytical insights (Patrizio, 2018). A commercial dataset, for instance, might allow estimation of traffic congestion by integrating weather conditions. The increasing adoption of IoT devices will lead to the generation of copious amounts of data. It is estimated that the global datasphere will increase substantially in size. Much of this data is heterogeneous and unstructured, resulting in challenges when attempting conventional analyses. In the next section, we will define big data harmonisation and discuss why organisations need it to gain actionable insights. In the absence of appropriate harmonisation techniques, an organisation could risk having plenty of data but little insight (Davenport, 2014)

The Big Data Harmonisation Concept and Its Significance

The procedure of aligning, standardising, and integrating data from different sources into a coherent dataset with a unified format and meaning is referred to as big data harmonisation. Disparate data is transformed into a "single version of truth" that can be examined analytically (Katal et al., 2013; Muhammad, 2025). This often entails merging datasets with regard to their historical context, cleaning and validating the data, reconciling units or terms discrepancies, and mapping disparate schemas into one canonical schema. In other words, data harmonisation creates reproducible processes concerning diverse datasets. Decades-old concepts in data warehousing such as Extract, Transform and Load (ETL) are still relevant today – though, because

of the colossal scope in our era's 'big data,' there is an unparalleled need for meshing disparate systems (Inmon, 2005; Kimball & Caserta, 2011).

Data harmonisation is often described as a prerequisite for organisations that wish to gain anything substantive from big data. By normalising formats and clarifying overlapping terminology, the process stifles inconsistency at its source and yields an analytics-ready dataset. Once the data is attended to in this way, dissimilar streams can be fused without the usual mismatch headaches, giving analysts a single canvas from which actionable insights emerge. An additional dividend is the lift that harmonised data gives to decision-making quality. When disparate sets are meshed correctly, predictive models benefit from a fuller picture, resulting in forecasts that command greater managerial confidence. Efficiency also improves because teams spend less time on cleanup and can redirect that time toward interpretation or strategy. Modern applications – including AI-driven content curation and instant fraud alerts – demand data sets that align flawlessly. Harmonised repositories let scientists launch machine-learning pipelines or visualisation dashboards without pausing to fix column names or bridge timestamp gaps, a chore that otherwise steals time from experimentation (IBM, 2023). This seamless infrastructure underwrites breakthrough uses such as risk forecasting in healthcare and individualised marketing nudges that respond to shifting user behaviours in real time. In brief, data harmonisation reconverts the disorder of raw input into a structured, trustworthy resource, opening the door for deeper insights and quicker action across research, operations, and strategies. When firms standardise and weave together data drawn from disparate streams, they protect the integrity of their numbers and still sharpen both decision-making and the performance of artificial-intelligence models. The resulting pool of harmonised big data turns into a powerful asset that, in everyday use, energises creative thinking, boosts productivity, and enables finely tuned customisation.

Applications in Daily Life Enabled by Big Data Harmonization

Big data harmonisation has fundamentally changed the way information is shared and exploited across disciplines as diverse as retail, medicine, urban transport, and streaming culture. By pooling and dissecting enormous, often heterogeneous volumes of incoming data, contemporary platforms can customise their offerings for single users while still scaling to millions of accounts. The world of online retail provides a conspicuous illustration of this trend. Companies such as Amazon, Shopee, and TikTok Shop harvest minute-by-minute logs of clicks, cart additions, and checkout completions, then shove that ocean of behaviour into recommendation engines designed to surface the products each shopper is most likely to love.

Another layer of complexity enters when marketers leverage the same unified datasets for customer segmentation. Activity recorded on a website, appended to a Customer Relationship Management (CRM) profile, and sometimes even observed in a brick-and-mortar store can be rapidly blended, after which algorithms scour the hybrid dataset and automatically generate messages like "you might adore these sneakers, since you picked up similar one's last month" (McKinsey, 2013). Those tailored nudges can appear as personalised subject lines in email, fresh hero images on a homepage, or even dynamic signage beside the cash register.

In recent years, mobile health- and activity-monitoring software has become almost ubiquitous, thanks in no small part to the popularity of wrist-worn fitness trackers and

smartwatches. Each of these devices records an array of personal data: daily steps, walking distance, caloric expenditure, resting heart rate, length and quality of sleep, exercise logs, food intake, and any weight fluctuations, thereby allowing the wearer to view nearly every facet of their lifestyle at one glance. A growing number of smartphone applications themselves, whether developed for iOS or Android, now pull and standardise this information through frameworks such as HealthKit or Google Fit, which let users see a unified snapshot even when the raw data originates from different gadgets.

By aggregating timestamps, location information, and biometric readings next to electronic medical files, the software has enough context to offer actionable recommendations customised for an individual. When, for instance, the system notices someone has neglected to meet a self-set step goal for five consecutive days, it may send a friendly reminder or nudge telling them to take a brief stroll, thereby turning raw numbers into tangible encouragement. Another recurrent practice in contemporary fitness software is the historical personalisation of exercise targets. An individual who declares an intention to shed pounds might be presented with a bespoke caloric deficit after the app surveys months or even years of logged steps, workouts, and food intake. Such a retrospective framing is demonstrably more productive than the cliche one-size-fits-all menu, chiefly because the advice feels timely rather than generic. The recommendation lives inside the user’s biography, quietly recalibrating as new entries arrive.

Population-scale datasets permit health tech firms to spot collective rhythms hidden beneath thousands of private journals. Engineers harvest signals that reveal when most people run, when they drift into light sleep by age group, and how chilly winter air nudges activity levels downward around mid-January. Those emergent patterns then redirect algorithmic engines, so community-wide challenges land in participants' inboxes precisely when they are most inclined to play. Wearables maintain the pulsing drumbeat of vigilance, chirping warnings when erratic heart behaviour suggests that a brief medical check may no longer be optional.
Incorporating personal health information into mobile technology routinely empowers individuals to make evidence-based lifestyle choices. Many contemporary fitness platforms assume the role of a behavioural coach, adapting workouts to the rhythms of daily life and personal growth. Analyses consistently demonstrate that feedback tailored to a user’s unique journey correlates with higher engagement, steadier exercise habits, and statistically significant gains in health markers. Parallels exist in the transportation sector, where applications such as Google Maps and Waze form the public face of big data integration. These services collate location signals, historical congestion patterns, and crowd-sourced incidents to recommend the quickest available route. The same datasets enable moment-to-moment re-routing suggestions that sidestep newly reported jams that would otherwise ensnare drivers. Such real-time adaptability emerges only when robust datasets are matched with automated vehicle controls that permit instantaneous route adjustments.

Researchers have long experimented with mash-ups of transit data to streamline the end-to-end journey. A rider might first drive to a suburban park-and-ride lot, then board a rail service whose timetable is mixed with updated road and platform conditions. An app such as Moovit – or even the more familiar Google Maps – will display the next bus, live tracks, and the odds of missing a connection, because each draw from GPS beacons, dispatcher feeds, and third-party partners that most users never see. The big data overlay alters everyday commuting in plain sight. Smartphones coach drivers

around trouble spots and shave minutes off fuel bills, which offers immediate gratification from the user. When a critical mass of motorists follows that guidance, the ripple effect can unclog arterial streets, even when the original road footprint remains fixed. Planners also benefit: anonymised bid-data packets that Google dispatches back to municipal traffic managers highlight persistent bottlenecks, nudging engineers toward timely interventions.

App-based taxi and courier platforms frequently rely on big data harmonisation to operate effectively. Companies such as Uber and Grab continuously ingest location pings from their drivers, incoming requests from passengers, live traffic feeds, and user star ratings in order to gauge the best possible driver-passenger pairing at any given moment. By blending that geospatial output with stored customer profiles and earlier trip histories, dispatch systems can refine routing, trim average wait times, and balance supply-and-demand pressures. Put simply, the transportation sector's embrace of these data integrations has given rise to what is often called smart mobility, an arrangement that empowers travellers to reach their destinations more directly while prompting platforms to respond to demand spikes with unusual agility.
Streaming platforms have become the public face of big data in daily life. Netflix, YouTube, and Spotify quietly harvest what you watch, how long you linger, and even what you skip in order to nudge a fresh recommendation onto your screen. Netflix, for example, quietly logs every pause and rewind, every click that hovers just long enough to light up a title card. Those individual moves are stitched together with genre tags, actor lists, and thematic keywords already attached to the film or show. The system then compares your profile with the profiles of thousands of viewers who share similar habits – a look-alike clusters – and arrives at the next watch that feels both novel and eerily familiar.

In addition, Netflix often displays multiple thumbnail images for the same film, and which one appears first can depend on whether you usually queue light-hearted rom-coms or weightier period dramas. The company credits these small nudges to its analytics engine, estimating that automated suggestions, not user searches, account for roughly 75 to 80 percent of what people wind up watching. Many music platforms repeat the playbook in sound. Spotify and Apple Music, for instance, log every track you play, skip, or save, then flag the time of day and playlist genre to predict your next listen. Meanwhile, YouTube composes its autoplay line-up from a symphony of personal and collective data, weighing Terms of Service, likes, dislikes, watch time, and the viewing habits of millions in real time. The next video that loads is not random; it reflects a global memory bank that has, for better or worse, learned exactly what tends to hold a specific account in place. TikTok's For You feed operates under much the same logic; whenever the user pauses, swipes, or rewinds, those tiny signals are swept into an internal model that assembles the following clip with almost frantic precision.

Issues and Suggestions

Large-scale data-gathering enterprises constantly stumble over the absence of an agreed-upon framework, a shortfall that forces analysts to wrestle with the messy job of aligning widely scattered information. In everyday practice, mismatched schemas, divergent semantic interpretations, and tightening ethical mandates stack the odds against seamless integration. Experts recommend starting with a ranked domain hierarchy that lays out a single logical model for the material in hand because doing

so creates obvious way-points for capture controls and clarifies the paths between raw source records and the destination dataset.
Artificial-intelligence routines provide another line of attack by dynamically adjusting dependency allowances wherever fields go missing, a trick that adds surprising flexibility and broadens usability. Between any two systems, quick-turn statistical recipes, machine-learned retrievals, and self-tuning semantic frames can form a scaffolding that keeps processing afloat. Tightening practical uniformity also hinges on small details: everyone gains when date-time stamps, measurement units, and code vocabularies settle into agreed formats. Pairing those habits with a living data dictionary and a sturdy Master Data Management layer gives routine inconsistencies a much smaller target. Finally, bringing all original data authors into the conversation up front and asking them to follow shared rules can slash the variety of problems originating before the data ever leaves their source.

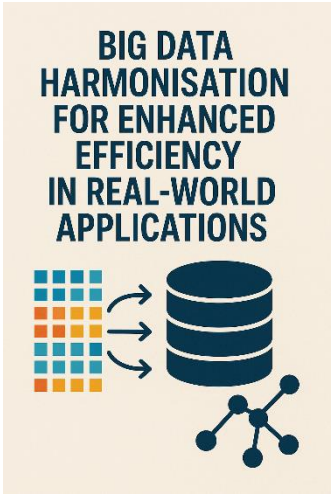
Quality-related data issues

Large datasets are often plagued by hidden quality problems, such as typos, duplicates, and rogue outliers. Even a few bad entries can trick experts into drawing misleading conclusions. One way to catch these lapses up front is to build multi-step pipelines that automatically harmonise and validate the incoming data. Rule-based flags, including regular-expression checks, will signal empty fields or unexpected formats. Machine-learning routines can add a second layer by spotlighting values that simply don't belong. Documenting where each data set comes from and what paired with a distributed Not only Structured Query Language (NoSQL) store on the cloud. That elastic architecture lets teams beef up processing power during traffic spikes and dial it back when the rush fades. In a widely distributed environment, mismatched copies of the same record can snowball into full-blown consistency crises. Techniques like two-phase commit, idempotent updates, and aggressive retry handling act as safety nets. Investing in scalable integration layers that absorb incoming volumes smoothly pays dividends in long-term stability.

Anytime identifiers cross systems, the risk of leaking personal information rises. Harmonising transformations it has undergone allows analysts to weed out any tainted batches. In practice, keeping validation thresholds a bit loose tends to raise the overall reliability of reported numbers.

Real-time data streams pour in at odd hours and in unpredictable volumes, so working with them can chew through hardware and network limits quickly. The classic ETL grind slows to a crawl when update windows shrink to seconds instead of hours. Frameworks such as Hadoop MapReduce or Apache Spark help, especially when
schemas and aggregating profiles may expose more than analysts intended, sparking both legal and ethical alarms. Strong encryption and strict access controls must run alongside the technical measures to protect the individuals behind the numbers. Laws such as the General Data Protection Regulation (GDPR) in Europe and Health Insurance Portability and Accountability Act (HIPAA) in the United States impose tight restrictions on how organisations collect, store, and share personal data. To remain compliant, many firms now bulk-up governance policies, lock down user permissions, and use encryption to mask data while it is in use. Before any two databases are joined, privacy-impact assessments force architects to

confirm that existing consents still cover the new purpose. Layering data, anonymising names, and keeping detailed logs of every transformation step help create an auditable trail, while federated analysis lets researchers compare algorithms without moving the original data out of its home system.



Many organisations get stuck because departmental silos keep engineers, analysts, and compliance officers from speaking the same vocabulary. Training programmes that boost digital literacy and short-term hires who build prototype pipelines can break that logjam. Senior stewardship teams then have to rally the groups around a shared vision, push user-friendly tooling, and keep technical jargon to a minimum. Giving everyone permission to treat information through a single corporate lens, even staff with limited SQL experience, reduces the friction that often slows projects. Data landscapes do not remain still for long; fresh feeds, updated schemas, and changed Application Programming Interface (APIs) arrive in weeks, even days.

Institutions committed to responsive data stewardship tend to draft explicit maintenance protocols paired with visual release notes that signal consumers about schema updates. Automated daemons then sift through each data river, comparing live flows against established baselines and sounding alerts when stray patterns disrupt normalcy. Modular architectures let developers bolt on new features without gutting older components, yet the ground rules that unite these modules – namely, those governing quality, privacy, and security – demand regular reassessment to stay relevant. When these practices align, disparate data sets can pool in a single store with surprisingly little human intervention, clearing the way for robust analytics, machine-learning experiment, and the day-to-day decision-making that keeps the enterprise nimble.

Conclusion

As previously mentioned, the processes necessary for big data integration are restructuring disparate, unorganized big datasets into useful information. Integration and synchronisation of separate datasets is crucial in order for different organisations to gain recognition through improved dealings with customers. These advancements are supported by AI systems extended over the Internet, with social media data helping to refine enterprises' technologies. When incorporated into a business, better data-driven insight leads to better decisions that allow innovation to flourish, which will eventually give a competitive edge when harnessed properly. It can allow a huge shift in added value from one business to many other participants, enabling the appropriate allocation of economic resources. This also impacts overall marketing operations, thus optimally rendering services where they are most needed and useful without value distortion. This process refines customer satisfaction through faster billing, a better gauge on usage pattern and the establishment of integrated responsive systems

This paper examines the persistent obstacles surrounding big data harmonisation, focusing in particular on data quality, system scalability, and privacy. It argues that uniform data standards, paired with robust ETL processes and machine-learning

techniques, can cleanse and merge disparate datasets; strong governance frameworks, backed by advanced security measures, are equally vital to safeguarding sensitive information. Meeting these hurdles is not an afterthought; organisations must allocate resources up front to tools, qualified personnel, and clear policy roadmaps if they hope to capture enduring value. Personalisation sits at the heart of any responsive application, allowing systems to ingest real-world signals and adjust behaviour almost in real time. As new waves of technology – the Internet of Things, smart cities, and generative artificial intelligence – pour ever larger volumes of data into the mix, the case for harmonisation grows increasingly clear. Seamless integration enriches everyday life, offering context-aware nudges, instant insights, and predictions that feel both timely and relevant. When executed with care, the harmonisation project does more than streamline workflows; it becomes an engine for innovation, unlocking latent value while still honouring privacy and quality standards.

References

Eaves, D. (2025). *Global digital usage statistics 2024: What happens every internet minute?* Retrieved from https://shithot.co.uk

IBM. (2023). *Why data harmonization matters in machine learning pipelines*. Retrieved from https://www.ibm.com

IDC. (2021). *DataSphere and StorageSphere forecast, 2021–2025*. International Data Corporation. Retrieved from https://www.idc.com

Inmon, W. H. (2005). *Building the data warehouse* (4th ed.). Wiley.

Davenport, T. H. (2014). *Big data at work: Dispelling the myths, uncovering the opportunities*. Harvard Business Review Press.

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: Issues, challenges, tools and good practices. In *2013 Sixth International Conference on Contemporary Computing (IC3)* (pp. 404–409). IEEE. https://doi.org/10.1109/IC3.2013.6612229

Khan, N. A., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud, A. S., Alnumay, W., ... & Gani, A. (2019). Big data: Survey, technologies, opportunities, and challenges. *The Scientific World Journal,* 2019. https://doi.org/10.1155/2019/7123761

Kimball, R., & Caserta, J. (2011). *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. Wiley.

Marr, B. (2016). *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Result*s. Wiley.

McKinsey & Company. (2013). *Big data: The next frontier for innovation, competition, and productivity*. Retrieved from https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation

Muhammad, H. (2025). Trends and projections in Big Data governance and harmonization. *Journal of Data Management, 19*(1), 33–48.

Patrizio, A. (2018). What is big data? More variety, volume and velocity. *Network World*. Retrieved from https://www.networkworld.com/article/3268073/what-is-big-data.html

SAP. (2023). From raw to refined: The value of big data harmonization. *SAP Insights*. Retrieved from https://www.sap.com/insights/big-data-harmonization.html

Shafer, J. (2017). *The rise of 42 Vs of big data.* Retrieved from https://www.datasciencecentral.com/profiles/blogs/the-42-v-s-of-big-data-and-data-science