IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 6, Issue 3, January 2026

Impact Factor: 7.67

A Survey on Multi-Agent LLM Frameworks for Data-Driven Dashboard Automation and Insight Generation

Dr. Riyazahemed Jamadar[1], Chaitanya Thakre[2], Shreya Kulkarni[3], Vaishnav Kokate[4]

Department of AI & Data Science

AISSMS Institute of Information Technology, Pune, India

riyaz.jamadar@aissmsioit.org, chaitanyathakre13@gmail.com

shreya.kulkarni132@gmail.com, vaishnavkokate15@gmail.com

Abstract: *The growth of enterprise analytics has increased the need for automated systems that turn raw, diverse data into clear insights and dashboards for decision-making. Traditional business intelligence workflows depend on manual interpretation, modeling, and visualization. This leads to slow and inconsistent analysis that relies heavily on expert knowledge. Recent developments in Large Language Models (LLMs) have improved the understanding of structured data and narrative explanations; however, using a single model is still not enough for complex analysis and understanding specific domains. Multi-agent LLM frameworks solve this issue by sharing analytical tasks among specialized agents. These agents focus on data profiling, domain knowledge, reasoning for insights, reflective improvement, and planning for visualization. This survey gives a detailed overview of multi-agent LLM designs for automated data-to-dashboard processes. We create a classification of system components, review key frameworks, assess performance trade-offs, and highlight emerging challenges like controlling inaccuracies, ensuring evaluation reliability, providing clear explanations, and addressing governance issues. Finally, we suggest open research opportunities for developing strong, scalable, and professional-grade autonomous analytics systems*

Keywords: Survey, Multi-Agent LLMs, Enterprise Analytics, Insight Generation, Dashboard Automation, Tree-of-Thought, Reflexive Reasoning, G-Evaluation, Data Interpretation

I. INTRODUCTION

In modern businesses, making decisions based on data is essential for operational efficiency, strategic planning, and gaining a competitive edge. Companies constantly produce large amounts of structured and unstructured data from their processes, digital interactions, customer behaviours, and operational metrics. However, turning this raw data into useful insights and dashboards depends heavily on data analysts and subject matter experts. Analysts need to understand distributions, correlations, hidden patterns, and key business indicators to create stories that help leaders make informed decisions. This process is not only resource-intensive but also faces challenges like skill gaps, cognitive biases, inconsistent interpretations, and limits on scalability that come with human-driven analytics. [1]. Traditional Business Intelligence (BI) platforms like Tableau, Qlik, and Power BI offer visualization and reporting tools, but they do not automatically reason or create narratives. They depend on users to know what to analyze, which metrics are important, and how to interpret insights in a specific context. As organizations grow, the variety of business contexts and analysis needs increases. This makes creating insights manually less efficient and harder to standardize across teams and decision-making levels. Large Language Models (LLMs) have shown strong skills in reasoning, summarization, creating analogies, and providing structured explanations. This comes from their ability to encode complex semantic knowledge. Techniques like Chain-of-Thought (CoT) prompting [2], self-consistency reasoning [3], Tree-of-Thought structured exploration [4], and reflexive self-correction loops [5] have further improved their interpretive depth. However, a single LLM operating as a standalone agent often struggles to reliably perform multi-stage analytics

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, January 2026
Impact Factor: 7.67

workflows. Insight generation needs several cognitive skills: data understanding, domain grounding, reasoning, narrative synthesis, and visualization design. When these roles are combined into one model, the insights are often shallow, too generic, or fabricated. To address these limitations, recent research has shifted toward multi-agent LLM architectures, in which several specialized LLM-based agents collaborate to emulate the workflow of a human analytical team [6], [7]. Each agent has a specific role, like interpreting datasets, detecting business concepts, deriving metrics, reasoning about causes, synthesizing insights, refining critiques, or planning visualizations. Coordination among agents enables reasoning to happen in cycles. Insights can be assessed, questioned, and improved before the final presentation. A key milestone in this direction is the Data-to-Dashboard framework introduced by Zhang and Elhamod [8], which demonstrated that insight quality improves significantly when analytic reasoning is decomposed into a structured agent workflow backed by reflective improvement cycles. Subsequent evaluation frameworks such as G-Eval [9] and HELM [10] have further supported the objective assessment of insight coherence and correctness. Yet, challenges remain in hallucination mitigation, domain alignment, visualization correctness, interpretability, and governance compliance [11], [12]. This survey reviews the current research on multi-agent LLM systems that are used for generating analytic insights and automating dashboards. We provide a classification of the architectural components, compare notable frameworks, discuss their strengths and weaknesses, and point out research areas that need further exploration to improve scalable, understandable, and business-ready autonomous analytics systems.

II. BACKGROUND AND FOUNDATIONS

The rise of Large Language Models (LLMs) has changed natural language systems from simple text generators to general-purpose reasoning tools. They can perform tasks like abstraction, summarization, classification, and structured interpretation. Their ability to understand meanings and create clear explanations makes them useful for analytical tasks that usually need experienced human analysts. However, to get reliable and relevant analytical results, we must understand how LLMs reason, how they work with structured data, and how they react to external sources like retrieval or rule-based limits. This section covers these basic elements.

A. Large Language Models as Analytical Reasoners

Modern LLM architectures such as GPT-4 [13], LLaMA [14], and mistral is trained on large-scale datasets that capture linguistic, statistical, and conceptual relationships. Its reasoning ability develops naturally rather than being programmed directly. This means the models learn to make multi-step inferences by being exposed to patterns in the data. Several prompting strategies have been shown to significantly enhance reasoning performance:

- **Chain-of-Thought (CoT):** Enables explicit stepwise reasoning by prompting the model to articulate intermediate logical states [2].
- **Self-consistency:** Uses multiple generated reasoning paths and selects the most convergent answer to reduce logical variance [3].
- **Tree-of-Thought (ToT):** Explores multiple reasoning branches and evaluates them iteratively for deeper reasoning depth [4].
- **Reflexive Reasoning:** Allows the model to critique and correct its own output, functioning as an internal evaluator [5].

These strategies are essential because multi-step data insights require explanation, interpretation, and justification—not simply answer generation.

B. Domain Grounding and Retrieval Mechanisms

Analytics workflows need a specific understanding of metrics, KPIs, aggregation hierarchies, and business rules. However, LLMs do not have lasting or certain knowledge of an organization's context. This can result in hallucination or misinterpretation when the internal business language differs from general language usage.

To address this, Retrieval-Augmented Generation (RAG) systems [15] augment LLM reasoning with authoritative domain sources such as:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, January 2026
Impact Factor: 7.67

- Enterprise knowledge graphs
- ERP data dictionaries
- Metadata Catalogs
- Standard KPI definition repositories

Grounding ensures that insights are tied to accurate organizational semantics rather than generic model reasoning.

C. Visualization Semantics and Narrative Alignment

Business dashboards must not only display data but must express the intended analytical story. Prior visualization intelligence approaches fall into three primary categories:

Rule-based visualization systems such as Draco encode hard constraints on chart selection based on data type and task [16].

Grammar-based frameworks such as Vega-Lite and Voyager enable expressive and compositional chart specification [17].

Machine learning-driven visualization recommendation systems (e.g., VizML [18], Data2Vis [19]) predict chart types based on learned patterns in data and prior visualization examples.

However, these systems historically lacked the ability to explain why a visualization is appropriate or how it supports a narrative insight. Multi-agent LLM visualization planners address this gap by linking statistical properties (e.g., variability, seasonality, comparison) with communication intent (e.g., trend explanation, comparison emphasis, anomaly justification).

D. Need for Multi-Agent Coordination

While single LLMs can handle specific reasoning tasks, the cognitive load for complete analytics pipelines (interpretation, reasoning, validation, visualization) is more than what a single prompting session can reliably manage. Multi-agent systems share the analytical responsibility, allowing for:

- Role specialization for deeper correctness
- Iterative critique-based refinement to reduce hallucination
- Decomposition of analytical reasoning into interpretable stages

This shift forms the foundation of emerging Data-to-Dashboard automation pipelines.

III. TAXONOMY OF MULTI-AGENT DATA-TO-DASHBOARD ARCHITECTURES

Multi-agent data analytics systems are networks of specialized reasoning agents that work together. They mimic the step-by-step reasoning and checking process that human analysts typically use. Although these systems can differ in their designs, many share a similar pattern of gradually improving analysis. To help analyze these systems, we suggest a framework based on six key roles that agents need to generate reliable insights.

TABLE I: TAXONOMY OF MULTI-AGENT LLM ANALYTICS SYSTEMS

System Component	Primary Role
Data Profiler	Data type inference, descriptive statistics, schema extraction
Domain Detector	Business domain recognition and KPI contextualization
Concept Extractor	Identification of metrics, measures, drivers, and causal features
Insight Generator	Descriptive, diagnostic, predictive, and causal reasoning
Self-Reflector	Critique, correction, uncertainty reduction, risk scoring
Visualization Planner	Chart grammar selection and narrative alignment

Data Profiler

The Data Profiler agent assesses the dataset by looking at column types, missing values, value distributions, segmentation patterns, and correlation signatures. This step makes sure the system starts with a correct statistical

ISSN: 2581-9429
Volume 6, Issue 3, January 2026
Impact Factor: 7.67

understanding of the raw data structure. Without profiling, later analysis risks making wrong assumptions about categorical and numerical variables, time series periodicity, or scale normalization.

Domain Detector

Enterprise metrics gain meaning from the context of their domain. For example, ”retention” in marketing has a different meaning than ”uptime” in DevOps. The Domain Detector connects dataset meanings to relevant business interpretations by using retrieval grounding and controlled vocabulary matching. This process greatly reduces mistakes by linking interpretation to established domain language.

Concept Extractor

This agent identifies key operational concepts like revenue drivers, churn indicators, cost centers, and performance KPIs. Instead of just observing patterns, the system must understand which attributes matter. Concept extraction connects statistical variation with operational meaning.

Insight Generator

The Insight Generator produces actual business insights through multiple layers of reasoning:

- **Descriptive:** What is happening?
- **Diagnostic:** Why is it happening?
- **Predictive:** What is likely to happen next?
- **Causal/Counterfactual:** What would change the outcome?

Tree-of-Thought exploration enables hypothesis branching, while Self-Consistency ensures insight stability across probabilistic reasoning paths.

Self-Reflector

The Self-Reflector acts like a senior analyst reviewing a junior analyst’s report. It checks:

- Logical coherence
- Statistical alignment
- Domain correctness
- Communication clarity

This step significantly improves reliability by identifying unsupported claims or illogical correlations.

Visualization Planner

The Visualization Planner translates analytical insights into clear visual representations that convey the intended message to decision-makers. Unlike traditional visualization recommendation systems that mainly depend on data type rules, this agent considers both what the data means and why the insight is important.

The planner first interprets the narrative focus of the generated insight. It determines whether the goal is to highlight a trend, compare groups, reveal a distribution, or identify an anomaly. Based on this intention, it chooses a suitable visual encoding strategy. For example, time-evolving performance indicators are best shown with line charts. Category comparisons work well with bar charts, while proportional breakdowns can use stacked or pie charts

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, January 2026
Impact Factor: 7.67

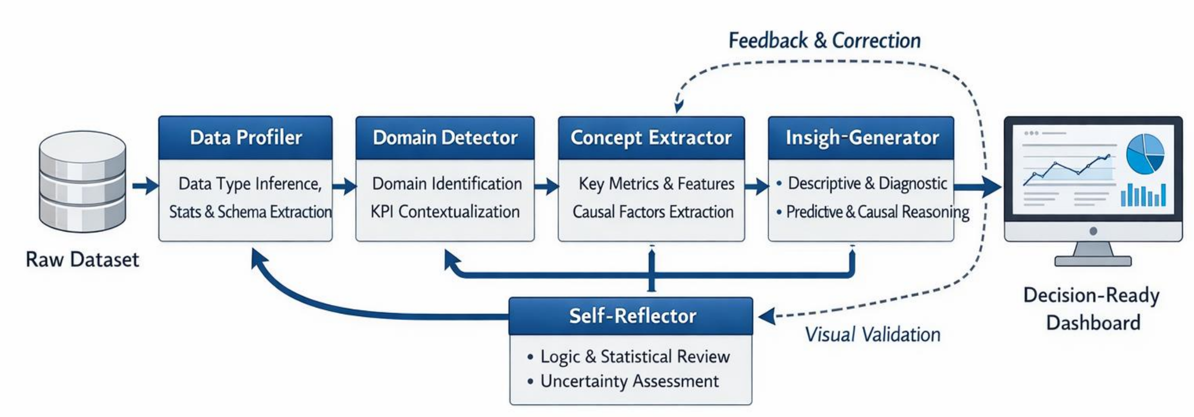

Figure 1: System Architecture of the Multi-Agent Data-to-Dashboard Analytics Framework

To ensure representational accuracy, the agent evaluates the following factors:

- **Data Characteristics:** Variable type (categorical, numerical, ordinal), value distribution, and presence of outliers.
- **Analytical Intent:** Whether the objective is explanation, comparison, forecasting, or anomaly emphasis.
- **Cognitive Load and Interpretability:** Preference for clean, minimal, and easily interpretable chart structures to avoid overwhelming users.
- **Domain Conventions:** Recognition of visualization formats commonly used within the specific business or industry context (e.g., funnel charts in marketing pipelines, waterfall charts in financial reporting).

Additionally, the Visualization Planner can create natural language annotations, titles, and callouts that direct users to important insights. This connects the visuals with the narrative flow instead of relying only on the end-user for interpretation. In some cases, the planner checks visual choices by working with the Self-Reflector agent, which allows for improvements when charts do not match the intended analysis. This teamwork makes sure the created dashboards are not just visually clear but also meaningful and focused on decision-making.

IV. COMPARATIVE SYSTEM ANALYSIS

Multi-agent analytics systems improve interpretability and insight depth, but require coordinated reasoning. Table II compares representative system classes.

TABLE II: COMPARISON OF SYSTEM CAPABILITIES

System	Domain Awareness	Reasoning Depth	Visualization Support
Single Prompt LLM	Low	Low	None
Business QA Bots	Medium	Medium	None
Multi-Agent CoT Pipelines	High	High	Partial
Data-to-Dashboard Systems	Very High	Very High	Strong

Single-agent LLM prompting often leads to shallow summaries that lack actionable significance. Multi-agent CoT systems provide deeper interpretation, but they may lack consistent visualization. Fully structured Data-to-Dashboard pipelines achieve a clear narrative and justify insights by coordinating analysis, reflection, and visualization components.

V. RESEARCH TRENDS (2021–2025)

The research area of multi-agent LLM-based analytics has changed quickly in recent years. This change is due to improvements in foundation model abilities, reasoning methods, visualization rules, and the needs of enterprise data management. The main trends in building automated insight generation systems are listed below.

A. Shift Toward Multi-Step and Structured Reasoning

Early applications of LLMs in analytics were limited to descriptive summaries of datasets, offering surface-level observations without causal interpretation. With the development of Chain-of-Thought prompting [2] and self-consistency reasoning [3], models began generating multi-step inference traces that resemble analytical thinking. This trend accelerated with Tree-of-Thought frameworks [4], where models explore different reasoning paths before selecting the best insight narratives. The shift from single-turn responses to deliberate reasoning workflows is essential for multi-agent architectures.

B. Rise of Agent Role Specialization and Delegated Cognition

The use of multiple interacting LLM agents reflects a shift toward delegated cognitive roles, where different agents specialize in data interpretation, domain conceptualization, counterfactual reasoning, insight refinement, and visualization design. Multi-agent orchestration frameworks such as AutoGen [6] demonstrate that decomposing complex reasoning into modular, purpose-built agents significantly improves insight correctness, interpretability, and narrative coherence. The field is moving toward model collectives rather than single omnipotent models.

C. Integration of Retrieval-Augmented Grounding for Domain Alignment

Enterprise analytics needs a clear understanding of specific terms, KPIs, and how operations work. Retrieval-Augmented Generation (RAG) pipelines [15] and enterprise knowledge graphs are becoming more common in analytical agent workflows. They help ensure that insights match real-world business meaning. Without grounding, LLM reasoning can lead to domain drift and mistakes. Current research looks into adaptive retrieval loops where models check different knowledge sources based on their uncertainty in reasoning.

D. Visualization Planning as a Semantic Communication Task

The visualization layer has changed from basic chart recommendations, as seen in VizML [18] and Draco [16], to semantic visualization planning. Modern systems choose visual encodings not only by data type but also by narrative intent, cognitive load, and clarity of interpretation. This represents a shift from visualization rules to explainable visual storytelling. In this approach, charts are selected to enhance the understanding of insights instead of just showing the data.

E. Evaluation Frameworks for Insight Quality and Narrative Validity

Evaluating how insights are generated is still a major challenge. Correctness in analytics is not just about facts; it also involves interpretation. Recent evaluation systems like G-Eval [9] and HELM [10] offer scoring methods for reasoning coherence, faithfulness, and clarity. At the same time, Insight Bench [20] suggests a multi-step evaluation to assess how useful insights are in business settings. The current trend is moving toward multi-dimensional evaluation that combines semantic accuracy, domain relevance, narrative clarity, and visualization appropriateness.

F. Governance, Explainability, and Accountability in Decision Pipelines

As organizations think about using multi-agent LLM analytics for operational decision-making, questions come up about transparency, bias management, data privacy, and regulatory compliance. Research is increasingly focused on creating reasoning traces that are auditable and interpretable to meet enterprise governance standards [11], [12]. Instead of fully replacing human analysts, current trends support human-inthe-loop verification to ensure decision accountability.

VI. CHALLENGES

Despite clear advantages, multi-agent LLM systems face several unresolved challenges:

- **Hallucination When Domain Context is Weak:** Even with retrieval grounding, inconsistent terminology or sparsemetadata increases the risk of false insights.

- **Reasoning Traceability:** Deep reasoning chains are hard toaudit, making compliance difficult for enterprises.
- **Evaluation Ambiguity:** The quality of insights is subjective; automated scoring methods are still experimental.
- **Scalability Issues:** Multi-agent orchestration adds inference overhead, which is unsuitable for real-time reporting.
- **Visualization Validity**: Some systems reason correctly but choose visualization forms that distort interpretation.

VII. FUTURE WORK

Future research should focus on:

- **Causal Reasoning Models:** Moving beyond correlation to justify business impact.
- **Adaptive Domain Lexicons**: Continuous learning systems that dynamically update KPI terminology.
- **Enterprise Knowledge Graph Integration:** Strengthening grounding with semantic embeddings of internal data.
- **Transparent Reasoning Traces:** Creating audit-ready explanations to support compliance.
- **Low-Latency Multi-Agent Scheduling:** Optimizing inference pipelines for production use.

VIII. CONCLUSION

Multi-agent LLM frameworks mark a significant step forward in automated enterprise analytics. These systems break down the analytical workflow by using role-specific reasoning agents. This approach addresses major issues of singlemodel prompting, such as shallow interpretation, weak domain grounding, and a lack of alignment in visualization. By coordinating data profiling, domain understanding, concept extraction, multi-step reasoning, reflective self-correction, and visualization planning, these frameworks generate richer and more actionable insights.

Nonetheless, implementing these systems still presents challenges. Reliable domain grounding depends on high-quality knowledge sources. Insight reasoning must be verifiable and clear, and visualization choices should reflect the narrative intent rather than just data-type rules. Moreover, evaluating the quality of insights remains subjective, underlining the need for standard benchmarks and governance frameworks to ensure transparency, auditability, and responsible use in enterprise decision-making.

Looking ahead, research is likely to focus on better integration of knowledge graphs, neuro-symbolic reasoning, causal inference modeling, and low-latency coordination of distributed agents. The long-term goal is to create autonomous, self-correcting analytics pipelines that maintain clarity and trust while functioning at a large scale. As these systems develop, multi-agent LLM architectures could not only automate dashboard creation but also transform how organizations analyze data, plan operations, and make informed decisions.

REFERENCES

[1] R. Kitchin, The Data Revolution: Big Data, Analytics, and the Future of Decision-Making. SAGE Publications, 2014.

[2] J. Wei, X. Bosma, Y. Zhao, K. Guu, A. Yu, B. Lester, E. H. Chi, Z. Chen, P. Liu, and P. Liang, "Chain-of thought prompting elicits reasoning in large language models," in Advances in Neural Information Processing Systems (NeurIPS), 2022.

[3] X. Wang, J. Wei, D. Schuurmans, E. Chi, Q. Le, and D. Zhou, "Selfconsistency improves chain of thought reasoning in language models," arXiv preprint arXiv:2203.11171, 2023.

[4] S. Yao, D. Yu, J. Zhao, K. Narasimhan, and M. Chen, "Tree of thoughts: Deliberate problem solving with large language models," arXiv preprint arXiv:2305.10601, 2023.

[5] N. Shinn, M. Labash, and N. Gupta, "Reflexion: Language agents withverbal reinforcement learning," arXiv preprint arXiv:2303.11366, 2023.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, January 2026
Impact Factor: 7.67

[6] T. Wu, C. Qian, J. Zhang, P. Jain, S. Chen, Y. Fan et al., “Autogen:Enabling next-gen llm applications via multi agent conversation framework,” arXiv preprint arXiv:2308.08155, 2023.
[7] W. Zhao, D. Li, F. Chen, J. Zhang, and X. Huang, “A survey of largelanguage model-based agents,” arXiv preprint arXiv:2308.11432, 2023.
[8] R. Zhang and M. Elhamod, “Data-to-dashboard: Multi-agent llm framework for insightful visualization in enterprise analytics,” in Proceedingsof the ACM SIGKDD Conference on Knowledge Discovery and DataMining, 2025.
[9] Y. Liu, Y. Deng, S. Chang, and X. Xie, “G-eval: Nlg evaluation usinggpt-4 with better human alignment,” arXiv preprint arXiv:2303.16634,2023.
[10] V. Sanh, A. Webson, C. Raffel, A. Rush et al., “Helm: Holistic evaluationof language models,” arXiv preprint arXiv:2211.09110, 2022.
[11] A. Agarwal, M. Patel, and V. Shah, “Governance for ai-generateddashboards: Auditability and risk,” arXiv preprint arXiv:2403.01234,2024.
[12] N. Kim, J. Hwang, S. Lee, and J. Lim, “Measuring faithfulness in datastories,” Proceedings of the ACL, 2023.
[13] OpenAI, “Gpt-4 technical report,” arXiv preprint arXiv:2303.08774,2023.
[14] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei,N. Bashir, S. Bapna, Y. Batra, P. Bhargava et al., “Llama 2: Open foundation and fine-tuned chat models,” arXiv preprint arXiv:2307.09288,2023.
[15] L. Gao, Z. Dai, Y. Wang, Y. Chen, R. Zhang, L. Zettlemoyer et al.,“Retrieval-augmented generation for knowledge-intensive nlp: A survey,” arXiv preprint arXiv:2312.10997, 2023.
[16] D. Moritz, C. Wang, S. Haroz, R. Kosara, and A. Satyanarayan,“Formalizing visualization design knowledge as constraints: The dracosystem,” IEEE Transactions on Visualization and Computer Graphics,2019.
[17] K. Wongsuphasawat, D. Moritz, H. Park, B. Lee, and J. Heer, “Voyager2: Augmenting visual analysis with partial view specifications,” IEEETransactions on Visualization and Computer Graphics, 2017.
[18] K. Hu, I. Bakr, S. Li, and A. Satyanarayan, “Vizml: A machine learningapproach to visualization recommendation,” in Proceedings of the ACMCHI Conference, 2019.
[19] H. Sung, D. Kardas, J. Ng, and D. Smith, “Data2vis: Automaticgeneration of data visualizations using sequence-to-sequence models,”arXiv preprint arXiv:1904.03126, 2019.

DOI: 10.48175/IJARSCT-30945