



SDU

— 1996 —
UNIVERSITY

Journal of
Emerging
Technologies
and Computing
(JETC)



Journal of Emerging Technologies and Computing (JETC)

is a peer-reviewed, open-access national and international scientific journal. Thematic areas: Computer Science, Infocommunication Technologies, and Mathematics with Applied Aspects

Publisher: SDU University

The journal is registered and licensed as an online publication and printed journal by the Ministry of Information and Social Development of the Republic of Kazakhstan.

Certificate (Online publication): No KZ01VPY00120097

Certificate (Printed journal): No KZ71VPY00120098

ISSN: 3105-6342 (online)

ISSN: 3105-6334 (print)

Frequency: four times a year (March, June, September, December)

Website: <https://jetc.sdu.edu.kz>

Editor-in-Chief:

Dana Utebayeva, PhD, Assistant Professor, SDU University

Managing Editor:

Assem Talasbek, PhD, Assistant Professor, SDU University

Managing Editor for the "Mathematics with Applied Aspects"

Bayan Bekbolat, PhD, Assistant Professor, SDU University

Technical Editor:

Nurislam Assan, SDU University

Editorial Board

- **Shirali Kadyrov**, PhD, Associate Professor,
Ajman university (UAE)
ORCID: 0000-0002-8352-2597
- **Selcuk Cankurt**, PhD, Assistant Professor,
Vistula University (Poland)
ORCID: 0000-0003-0581-1913
- **Khaled Mohamad**, PhD, Assistant Professor,
SDU University (Kazakhstan)
ORCID: 0000-0002-5980-0147
- **Lyazzat Ilipbayeva**, Candidate of Technical Sciences, Associate Professor,
IITU (Kazakhstan)
ORCID: 0000-0002-4380-7344
- **Kamila Orynbeкова**, PhD,
SDU University (Kazakhstan)
ORCID: 0000-0002-2182-2914
- **Zhandos Dosbayev**, PhD,
Satbayev University (Kazakhstan)
ORCID: 0000-0003-1673-4036
- **Bektur Baizhanov**, Doctor of Physical and Mathematical Sciences, Academician, Professor,
SDU University (Kazakhstan)
ORCID: 0000-0002-3743-7404
- **Nurlan Dairbekov**, Doctor of Physical and Mathematical Sciences, Professor,
SDU University (Kazakhstan)
ORCID: 0000-0002-2725-7549

**Ministry of Science and Higher Education of the Republic of
Kazakhstan**

SDU University

**Journal of Emerging Technologies and
Computing (JETC)**

Volume 6, Issue 3 • September 2026

Kaskelen, Kazakhstan — 2026

CONTENTS

• SECTION I – Computer Science

- AI-Assistants: Framework for combining LLM-Driven Assistants with corporate features

Isakulov Zeiin 7

• SECTION II – Infocommunication Technologies

- Internet of Things for the Visually Impaired: Design of a Wearable Obstacle-Detection and Navigation Device

Marxuly Sunggat, Abdirashit Alua, Kalandarov Palvan, and Nemerbayeva Meruert 28

• SECTION III – Mathematics with Applied Aspects

- A Two-City, Two-Group SIR Model with Heterogeneous Risk-Perception Feedback and Group-Specific Mobility

Yerikkyzy Arailym 44

- Varieties of right-commutative algebras defined by identities $(aa)a=[a,b]a=0$

Abilakim Aidana and Zhakhayev Bekzat 55

- Space of Tortken elements in free Novikov algebra

Baigali Meruyert 63

SECTION I

Computer Science

This section focuses on current research directions and applied
advancements
in Computer Science, particularly in the areas of artificial
intelligence,
software engineering, and intelligent systems.

Article

AI-Assistants: Framework for combining LLM-Driven Assistants with corporate features

Isakulov Zeiin ^{1*}

¹Department of Information Systems, SDU University, Almaty, Kazakhstan

DOI: 10.47344/byvn3m48

Abstract

This paper presents a comprehensive review of existing studies and several optimization methods, and suggests a simplified workflow for a unified framework for multimodal systems and AI-driven assistants. It focuses on the integration of AI-based tools into existing IT infrastructures by comparing approaches used in industry use cases of applying AI-based tools in their respective fields. Integrating AI into existing IT infrastructures that were not originally designed to be operated by AI, even at the level of data rendering, causes significant challenges that are worth solving. The main challenges include evaluating the rationality of using AI for certain tasks, managing the computational costs of self-deployed AI tools, managing the quality limitations of self-deployed LLMs to ensure that they can effectively interact with humans in industrial environments, and overcoming the economic limitations of using high-end LLM models such as ChatGPT, DeepSeek, or Grok in terms of token-based pricing. The study demonstrates a significant need for standardization and enhanced operability of existing approaches beyond the highly specialized solutions used in fields such as medicine, software development, and network management. The improvements are intended to enable AI assistants to make critical decisions that still require automation and optimization within industrial systems. The main aim is to systematize modern solutions and design common principles and vectors of development for highly contextual and scalable AI operating systems, enabling both AI technology providers and commercial organizations to adapt even in the short term. An example of the workflow pipeline of the proposed framework is designed to combine the adaptability and intellectual operability of high-end LLMs with the standardization of tasks for highly contextual models. It also addresses the deployment limitations of models built from scratch or derived from local models while preserving optimization for servers of small and medium-sized enterprises at each step by managing small modular tasks.

Keywords: Artificial Intelligence, AI Adapters, Human-AI interaction, Industrial AI, Process Automation, IT infrastructure, multimodal systems, AI-driven assistants, cost-effective deployment, enterprise AI, scalable AI frameworks, SME AI adoption, computational cost management

Email*: isakul.qxz@gmail.com ORCID: 0009-0001-2930-5233

Received: December 20, 2025. Reviewed: September 26, 2026. Accepted: September 29, 2026. © 2026 Isakulov Zeiin. All rights reserved.

I. INTRODUCTION

Recently, AI-based workflows and ML algorithms have increasingly developed and become the basis for optimizing the work environment. With further growth in the volume of data created and recorded by humanity, and with growing computational requirements, traditional planning algorithms and task-completion methods require enhancement. The IT infrastructure of Big Tech companies, cloud systems, and distributed systems is increasingly using intelligent algorithms and AI-based tools, leaving traditional methods and tools, as well as small and medium-sized enterprises, far behind. The use of neural networks, AI assistants, and metaheuristic algorithms in the early period of their development opened a new chapter in achieving optimization, industrial effectiveness, and automation. These approaches allow us to understand the main benefit of technological progress—adaptability. Traditional solutions, despite their effectiveness in static environments, are not adaptable. The adaptability of a system is necessary for dynamic systems where data and rules are changeable, including large-scale cloud platforms, which are increasingly being adopted even by small and medium-sized enterprises in various fields.

Conventional IT management infrastructure is often dependent on static optimization algorithms and rule-based decision-making. Although this type of procedure is successful in stable environments, it often does not scale or respond to the unpredictability and volatility of today's IT ecosystems. Using different kinds of intelligent processing algorithms can help predict system bottlenecks in existing IT infrastructure, find ways to optimize workflows, and enhance our ability to achieve state-of-the-art adaptability. These algorithms can autonomously predict and identify system bottlenecks and then automatically reposition workflows for continuous operational efficiency, paving the way for adaptability.

It is also important to mention that the implementation of such intelligent tools may cause significant organizational changes, including employee re-education, and raise concerns such as data leaks and quality issues. The creation of artificial intelligence systems that meet both the technical constraints and the strategic investment requirements for algorithms in small and medium-sized enterprises remains an important task that requires a carefully developed approach. In this study, we will explore possible approaches.

II. LITERATURE REVIEW

Our review of previous studies gives us more context and helps us identify the main gaps, methods, and findings in the industrial field. The literature review examines both recent and older works, allowing us to delve deeper into the subject and helping us position our work on the more practical side of the research. The literature is organized into three thematic areas: optimization and scheduling approaches; studies on the industrialization of the RAG system, Part I; and studies on the industrialization of the RAG system, Part II.

TABLE I: Optimization and Scheduling Approaches

Study	Key Contributions	Rationale
[1]	<i>RepairLLaMA: Efficient representations and fine-tuned adapters for program repair</i>	Provides an example of using adapters and overall workflow pipelines.
[2]	<i>Exploring the potential of AI-driven optimization in enhancing network performance and efficiency</i>	Shows the efficiency of AI approaches in industrial cases.
[3]	<i>Software project scheduling using the hyper-cube ant colony optimization algorithm</i>	Provides evidence of heuristic approaches in industrial scheduling.
[4]	<i>An investigation of static versus dynamic scheduling</i>	Demonstrates differences between scheduling approaches.
[5]	<i>A roadmap for integrating automation with process optimization for AI-powered digital transformation</i>	Discusses the potential vectors of development and processes to be automated.

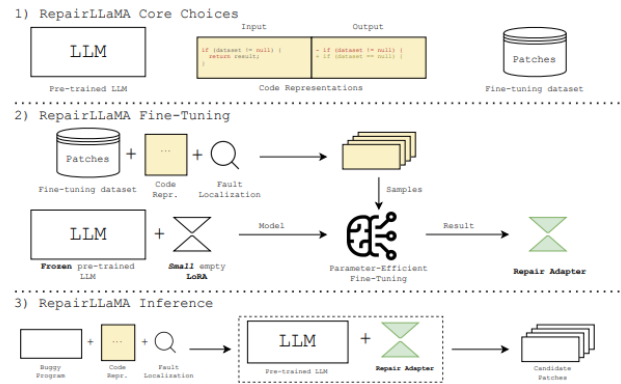


Figure 1: Overview of RepairLLaMA. The core novelties of RepairLLaMA are the APR specific code representations and the engineering of an effective *program repair adapter* that is plugged into the underlying LLM.

Fig. 1: The whole pipeline of the work [1]

The studies have shown the importance of optimization approaches in industrial cases. Early studies provided evidence of heuristic approaches in industrial scheduling [3]. Some of them demonstrate the differences between static and dynamic scheduling, proving the importance of adaptability [4]. However, with today's technological advances, studies discuss the vector of development toward automation, optimization, scheduling, and routing using intelligent tools in industrial environments [2]. Recently, RepairLLaMA demonstrated an application of fine-tuned adapters, enhancing both human-based and AI-based workflows in specific areas [1]. See Fig. 1.

TABLE II: Industrialization of RAG system(PART 1)

Study	Key Contributions	Rationale
[6]	<i>Model Context Protocol (MCP): Landscape, Security Threats, and Future Research Directions</i>	Primarily aimed at simplifying pipeline of deployment, integration and operation with existing technologies on an industrial scale. MCP offers an example of simplifying models workflow with each other in order to expand functionality with external tools and reduce fragmentation.
[7]	<i>What if GPT4 Became Autonomous: The Auto-GPT Project and Use Cases</i>	Provides use cases of autonomous AI and include potential vectors of development and enhancement of LLMs.
[8]	<i>Auto-GPT for Online Decision Making: Benchmarks and Additional Opinions</i>	Demonstrates the key differences across autonomous models and provides Additional Opinions algorithm which enables integrating smaller models without major changing the main LLM
[9]	<i>Financial Report Chunking for Effective Retrieval Augmented Generation</i>	The study shows significant improves in results of RAG in question and answer tasks based on financial tests. The approach allows us to create high-precision RAG systems using file structure instead of full texts.
[10]	<i>Searching for Best Practices in Retrieval-Augmented Generation</i>	The study examines existing RAG methods to identify best practices to ensure productivity and efficiency.

TABLE III: Industrialization of RAG system(PART 2)

Study	Key Contributions	Rationale
[11]	<i>Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks</i>	The study demonstrates RAG system model. RAG can solve problems of limited and accurate knowledge of LLM and the difficulty of updating them.
[12]	<i>When Will AI Exceed Human Performance? Evidence from AI Experts</i>	The study surveyed ML experts to shows that AI will surpass humankind in many areas in the coming years. Giving us potential vectors of implementing human-in-the-loop.
[13]	<i>Fundamentals of Expert Systems</i>	The study demonstrates key principles of creating expert systems and outlines areas for further research.
[14]	<i>Adaptable Adapters</i>	The study proposes adaptive adapters to select own activation functions for different layers and use a trainable switch to select only necessary layers. The approach preserves quality of standard adapters, yet requires less parameters for better for smaller data.
[15]	<i>Dressing the Imagination: A Dataset for AI-Powered Translation of Text into Fashion Outfits and A Novel KAN Adapter for Enhanced Feature Adaptation</i>	The study proposes NeRA, an innovative adapter architecture. Unlike traditional approaches, NeRA uses trainable nonlinear spline transformations enabling models to provide higher accuracy and deep semantic matching.

A significant part of the existing work is aimed at making AI assistants practical and contextual. The foundations of RAG were described and demonstrated to solve the knowledge accuracy and limitations problems of LLMs [11]. But as time goes on, the increase in the efficiency of AI assistants continues [10]. Concrete approaches to making models highly contextual include, for example, the capability to work with financial reports by using better chunking methods. To reduce the fragmentation of AI-based tools, the studies also proposed the Model Context Protocol as a basis for simplifying workflows and the integration of external tools. We can also see a tendency toward more effective and powerful customization of the models. The adaptable adapters can select functions to use and yet use fewer parameters and conduct quality for less volume of data [14]. Some approaches simply suggest architectural changes to the adapters to make models even more accurate by using different computational operations [15].

TABLE IV: Human-AI and AI-AI interaction

Study	Key Contributions	Rationale
[16]	<i>In-Context Retrieval-Augmented Language Models</i>	The study proposes In-Context RALM, a simpler method in which the architecture of the model does not change in comparison with regular RALM. The context documents are simply added to the beginning of input data. In-Context RALM has great potential for widespread use of the model through the API.
[17]	<i>Human-in-the-loop Artificial Intelligence</i>	The study proposes ethical concept of HitAI. Since humans are involved in any AI system, HitAI suggests rewarding those who provide knowledge. Profits from AI solutions should be returned to the rightful owners of knowledge.
[18]	<i>Towards a Better Understanding of Context and Context-Awareness</i>	The study addresses challenges related to understanding and creating context and suggests approaches of developing contextual applications.
[19]	<i>AI models collapse when trained on recursively generated data</i>	The study explains the risks of future versions of GPT. When most of the text on the Internet will be generated by LLM. It was found that uncontrolled generation of content by the model leads to irreversible defects of future models. The phenomenon is called "model collapse." It is found in many generative models.
[20]	<i>Artificial intelligence implementation in manufacturing SMEs: A resource orchestration approach</i>	The study examines how manufacturing SMEs organize resources to implement AI. Via dynamic orchestration of AI resources, SMEs successfully coordinate processes and develop qualified specialists. The research contributes to AI implementation by showing how SMEs use AI capabilities to create a competitive advantage.

Integration of AI into existing human-based infrastructures raises important considerations including ethical ones. *Human-in-the-loop Artificial Intelligence* suggests more ethical AI, the HitAI. The concept suggests that the profits from AI in enterprises should

be returned to their owners [17]. The long term risks like model collapse [19], and practical benefits [20] suggest to us that AI must be regulated to be the technology of the future. Some infrastructures were never meant to be driven by AI, but they still need some intelligent solutions. There, contextuality comes in the first place, both for human and AI agents [16] [18]. The reviewed studies demonstrate a lack of a standardized, lightweight framework and development pipeline. The studies present highly effective but specialized solutions [1] [2]. The existing research does not provide a solid platform that addresses computational constraints and implementation challenges faced by SMEs. The reviewed literature on autonomous AIs [8] [7], model interactions [6] and resource orchestration [20] hints at this but does not explicitly propose or test this specific hierarchical architecture as a virtual copy of existing corporate agents in SME infrastructures.

Case Studies: Google's use case of ML to optimize its data center cooling systems by analyzing data from sensors ensures that the system can dynamically manage server setup settings resulting in energy and financial savings. AT&T's ML algorithm analyzes patterns in its traffic and can predict demand, minimizing further network latency issues.

The integration of AI and ML algorithms in industrial cases demonstrates huge improvements in network management. Predictive analytics and anomaly detection in real time form the foundation for future AI-based corporate tools. It enables networks to react to current conditions and predict future threats.

The article's main limitation is that it was based on an industry leader's case. It doesn't present quantitative results from practical implementation of AI driven methodologies. It also lacks a detailed analysis of computational costs and complexity of integrating these systems into existing legacy infrastructure.

The article demonstrates effectiveness in energy savings and reduced latency and validates its potential. Its complexity is high, requiring interdisciplinary knowledge. It is a great example of ML optimization applied to load balancing and resource management.

After the literature review, it becomes obvious that enhancements have been made in the AI workflow pipeline. But there remains a substantial gap in the standardization of the frameworks used. That can be a limitation on fully implementing them in existing IT infrastructures. The research mainly focuses on specialized solutions without providing solid frameworks that address the computational constraints and implementation challenges faced by small and medium-sized enterprises.

III. RESEARCH METHODOLOGY

A. Research Questions and Objectives

Our research aims to contribute to the development of the standardization of frameworks used in the AI development pipeline by making them optimized even in small and strongly tied to context-based automation and workflow management in IT infrastructures.

This research addresses the following questions:

1. How can an LLM framework be successfully implemented within existing IT infrastructure to manage an AI assistant with significant computational constraints?
2. How does an AI architecture integrate classical optimization local models to optimize operations in industry with clients?

The aim is to design a framework for optimizing and scheduling tasks in multimodal systems that improves the efficiency and quality of combining models in IT infrastructure. Our aim also includes analyzing the implementation challenges and the organizational effects of intelligent automation. For this purpose, we have to face the following tasks: 1. To analyze the existing methods of implementing multimodal systems. 2. To develop a hybrid model combining LLMs and classical optimization algorithms. 3. To increase the effectiveness of the proposed framework under real-case scenarios.

B. Experimental Setup

The experiment was applied to an existing IT infrastructure environment and designed to recreate a real-world enterprise environment. The testbed consists of a digital infrastructure, storage systems, and API components required to test the system in a medium-scale enterprise. The testing system was built using HuggingFace models and OpenRouter API models to orchestrate under varying resource constraints and workload patterns.

The experimental setup includes five primary components tailored for the Unibest case study. The first one is a baseline RAG system implementing a multimodal structure. The second one is the proposed framework integrating several models and tools with the target AI assistant. The third is a management layer implemented as an orchestration UNIA positioned between AI assistants and a sales manager. The fourth one is the operational knowledge stores such as vector databases, relational product databases, rules, etc. The last component is monitoring, logging and CRM integration enabling more context for the system's work. The orchestration LLM acted as the leader of the "AI Swarm" that standardizes corporate tasks across smaller models by performing

prompt engineering and setting up constraints such as policies and rules. Each component was deployed in identical environmental conditions to ensure a fair comparison.

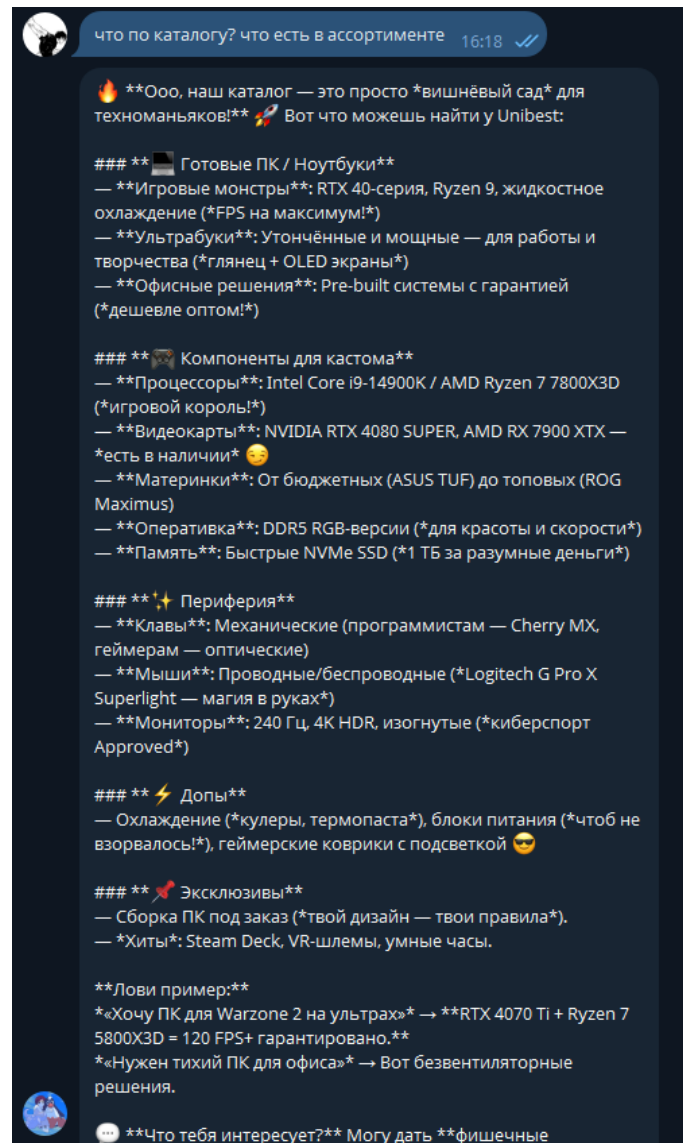
System deployment: The core inference model ran on a physical server with AMD Ryzen 7 9800X3D under Debian, an NVIDIA RTX3070 Ti GPU and 32 GB RAM. Copies of UnieAI used a single quantized Nous Hermes Mistral 7B model to reduce graphical memory and inference footprint. External channels were connected via APIs such as Telegram, WhatsApp, Instagram, Gmail. UNIA in the orchestration layer used DeepSeek via OpenRouter API with a constrained budget (50 free requests per day) for selected high-risk decisions and light-weight routing tasks such as managing prompts and controlling the "AI Swarm". Heavier inference and generation were performed locally on the quantized UnieAI.

C. System Architecture and Components

1) Unibest Business Context: Unibest is a B2B/B2C company selling computers and related services. To reduce human labor costs, a sales AI assistant, UnieAI, was created to act as a virtual sales manager handling customer communications, product recommendations, making store orders and routine after-sales interactions during non-working hours.

To manage and supervise UnieAI, a manager model named UNIA was implemented. UNIA is built atop the free DeepSeek endpoint available through OpenRouter (limited to 50 free calls per day) and provides orchestration features such as creation and lifecycle handling of UnieAI copies, routing of B2B leads, scheduling B2C review reminders, context recovery and quality correction of UnieAI outputs, especially for the Kazakh and Russian languages. UNIA can autonomously create and manage CRM objects in Bitrix, initiate procurement protocols executed by UnieAI and periodically update B2B data by using a 2GIS parser.

Knowledge and retrieval stack. The product catalog, company policies and trade methodology are stored in relational and document stores and mirrored into vector indices using FAISS and Qdrant. Embeddings for Russian and Kazakh were produced using an embedding model from Sberbank. RAG pipelines fetch the top-k passages from vector stores and pass those to UnieAI for grounded response generation.



D. Operational and Knowledge Update Pipelines

Product database changes propagate into both the canonical relational store and into the document ingestion pipeline. Documents are chunked, embedded using Sberbank embeddings for RU/KZ languages, and re-indexed into FAISS/Qdrant on a scheduled cadence or via delta-updates. See Fig.2 for an example of UnieAI using the products database.

Periodically, the system runs the 2GIS parser to enrich B2B data with newly found contacts and inserts them into the vector store and CRM as lead candidates for UNIA routing. See Fig.3 for a data table generated using the 2GIS parser.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	Наименование	Описание	Рубрика	Адрес	Координаты и Плотность	Метрополитен	Район	Полос	Одну	Решон	Страна	Часу работы	Часовой пояс	Рейтинг	Количество	Телефон 1	Телефон 2	Телефон 3	E-mail 1	E-mail 2	Веб-сайт 1	Веб-сайт 2	Instagram	Facebook	
2	BPO	Итого количество	Компьютерная	Адрес прощан	00002	Центральный	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,5	510	8703885741											
3	Бел Cyber Азия	Компьютерная	Компьютерная	улица Навои, 2	00043	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	5	353	8708500038											
4	LCD	Интернет-кафе	Компьютерная	проспект Навои, 1 этаж	00024	Мусорный рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,6	602	87423391											
5	MyGid.com	1 компания	Компьютерная	улица Рыбакова	00080	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	2,9	58	8747688408											
6	Валса	компьютерный	Компьютерная	проспект Тараса	00080	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,4	283	8702841041											
7	Qinet Cyber Club	компьютерный	Компьютерная	улица Тополева, 1 этаж	00048	Алматыский	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,4	371	8707760200											
8	Forum Cyber B	компьютерный	Компьютерная	улица Рыбакова, 3 этаж	00048	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,5	208	8749451273											
9	Cyber city	компьютерный	Компьютерная	улица Рыбакова, 102-0	00048	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	448	8707760217											
10	RAU803 Cyber	компьютерный	Компьютерная	улица Кавказа	00048	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	128	8771887388											
11	Lifeline	компьютерный	Компьютерная	проспект Дзюла	00010	Мусорный рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,7	347	874774488											
12	Diga Cyber Club	компьютерный	Компьютерная	улица Волжарского, 40а	00013	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	183	8707843887											
13	KIT	компьютерный	Компьютерная	Свободы	00042	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	5	4												
14	Qinet	компьютерный	Компьютерная	микрорайон Мн	00042	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,7	166	8700675077											
15	Jet Cyber Club	компьютерный	Компьютерная	улица Таубатаева, 87/2	00057	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	481	8707103011											
16	Cyber Citynet	компьютерный	Компьютерная	улица Дзюла	00012	Алматыский	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	631	8778014888											
17	The One Pro	услуг	компьютерная	Алматы-9	00051	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	281	8700109007											
18	Vinland	компьютерный	Компьютерная	улица Навои, 181	00042	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	25	8707689190											
19	Vinland	компьютерный	Компьютерная	улица Боровая, 10	00080	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	231	8708429874											
20	UWB	компьютерный	Компьютерная	улица Дзюла, 40	00038	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,4	166	8702218328											
21	Omni Cyber City	компьютерный	Компьютерная	проспект Магистраль, 172	00000	Мусорный рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	80	8747678171											
22	FI	компьютерный	Компьютерная	улица Муратаева, 2 этаж	00012	Алматыский	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	148	8707111020											
23	Thema Natus	компьютерный	Компьютерная	Ботанический	00012	Алматыский	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,7	82	8705232820											
24	Qinet Cyber Club	компьютерный	Компьютерная	Алматы	00048	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,7	119	8707603881											
25	Tatary Cyber Club	компьютерный	Компьютерная	улица Навои, 1 этаж	00038	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	358	8700687478	8700687484										
26	1405 Cyber Club	компьютерный	Компьютерная	микрорайон Тау	00042	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,7	235	8700111405											
27	MOHNET Cyber	компьютерный	Компьютерная	микрорайон Алатай-4	00048	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	313	8747818437											
28	Major Cyber Club	компьютерный	Компьютерная	улица Навои, 208	00043	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	287	8700689382	8700689381										
29	Omni Cyber Club	компьютерный	Компьютерная	улица Жукоса	00038	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	80	8747847388											
30	Diga Cyber Азия	компьютерный	Компьютерная	улица Тополева, 1 этаж	00012	Алматыский	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,7	27	8747807381											
31	Shantien	услуг	компьютерная	Байкалова	00048	Алматыский	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,7	104	8778183875											
32	Бел	компьютерный	Компьютерная	улица Жукоса	00080	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	30	8707103003											
33	IK Cyber Азия	компьютерный	Компьютерная	Е. Боровая	00048	Алматыский	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,7	224	8747308200											
34	Валса Cyber Club	компьютерный	Компьютерная	проспект Рыбакова, 101/1	00081	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	60	8707600847											
35	2803	компьютерный	Компьютерная	улица Кавказа	00048	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	186	8747602081											
36	Gama Nax X	услуг	компьютерная	улица Кавказа	00040	Алматыский	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,7	247	8747302081											
37	Волга Cyber	компьютерный	Компьютерная	микрорайон Алатай-1А	00051	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	138	8700689381	8771007485										
38	Алматы Cyber Natus	компьютерный	Компьютерная	улица Курмангазы, 141	00058	Алматыский	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,7	119	8700689381											
39	PAJF	услуг	компьютерная	улица Рыбакова, 111	00071	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	238	8707249000											
40	Валса	компьютерный	Компьютерная	улица Тополева, 101	00018	Мусорный рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	3,4	122	8702819003											
41	152777 Cyber	компьютерный	Компьютерная	проспект Садыр	00037	Трудовой рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	3,8	163	8771823388	8747177799										
42	Jet Cyber Азия	компьютерный	Компьютерная	микрорайон Мн	00042	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,7	82	8707388824											
43	Алматы Cyber	компьютерный	Компьютерная	улица Сулейманова, 20	00038	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,4	238	8747218888											
44	Cyber Net	компьютерный	Компьютерная	улица Жукоса	00000	Алматыский	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,4	355	8700707270											
45	Shantien Cyber	компьютерный	Компьютерная	улица Алатай-4А	00008	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	141	8708678007											
46	DIG Cyber Club	компьютерный	Компьютерная	улица Байкалова, 2 этаж	00058	Трудовой рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	80	8700689381											
47	Наименование	компьютерный	Компьютерная	улица Боровая, 1018	00018	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	8	1000	8771278212											
48	Leid	компьютерный	Компьютерная	улица Магистраль, 2 этаж	00000	Алматыский	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	783	8708302222											
49	PAJF	услуг	компьютерная	улица Жукоса	00008	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,2	148	8707738884											
50	Proad	компьютерный	Компьютерная	улица Кавказа	00008	Алматыский рай	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,7	78	8778838000											
51	PAJF	услуг	компьютерная	улица Тополева	00040	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	4,8	161	8774400388											
52	Валса	компьютерный	Компьютерная	улица Магистраль, 11 этаж	00081	Ботанический	Алматы	Алматы	Казахстан	Пн 00:00-24:00	+05:00	3,8	47	8700430000											

Fig. 3: Example of product table based on "comp clubs" search in Almaty

UnieAI prompt templates and small fine-tuned adapters are versioned, meaning that UNIA can push prompt updates to running copies and trigger cache invalidation of previous contexts. See Fig. 4 for the overall architecture illustration.

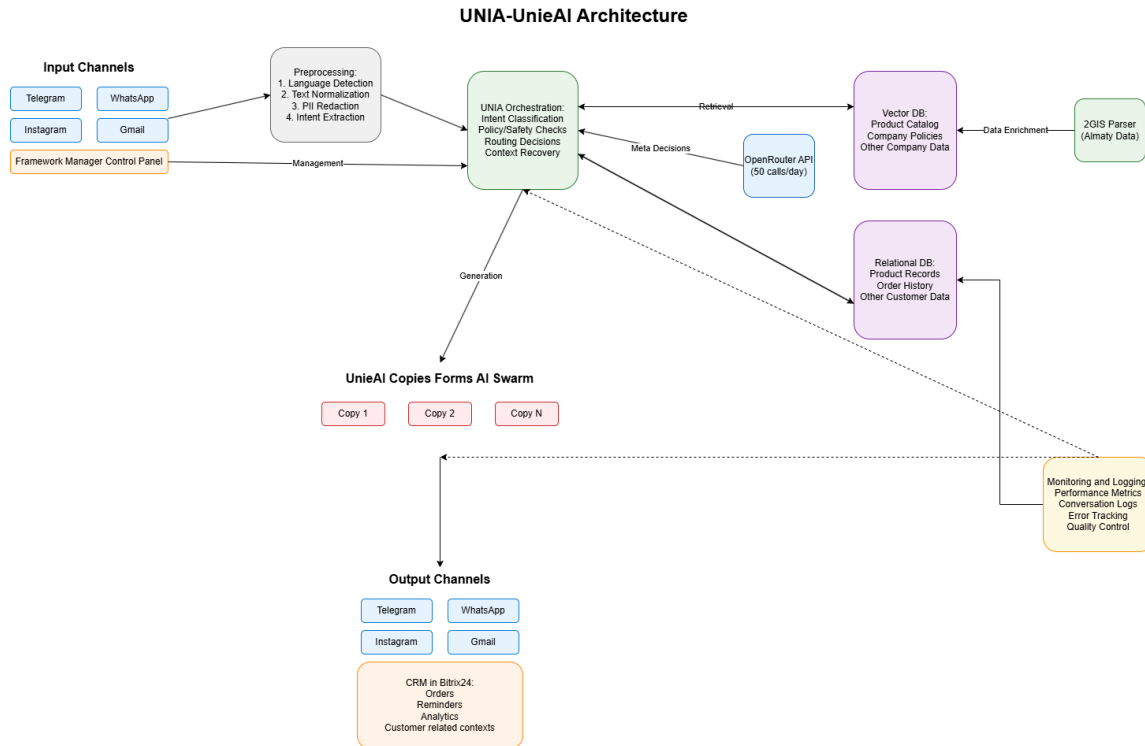


Fig. 4: Algorithm

E. Operational Pipeline — Client Input

The end-to-end pipeline for a customer message is as follows:

- 1) Customer sends a message via Telegram/WhatsApp/Instagram/Gmail. The channel adapter normalizes incoming payloads and forwards the message to the preprocessing layer.
- 2) Language detection, text normalization, adapting markups and formatting and intent-slot extraction via lightweight embedding and translating models using masking.

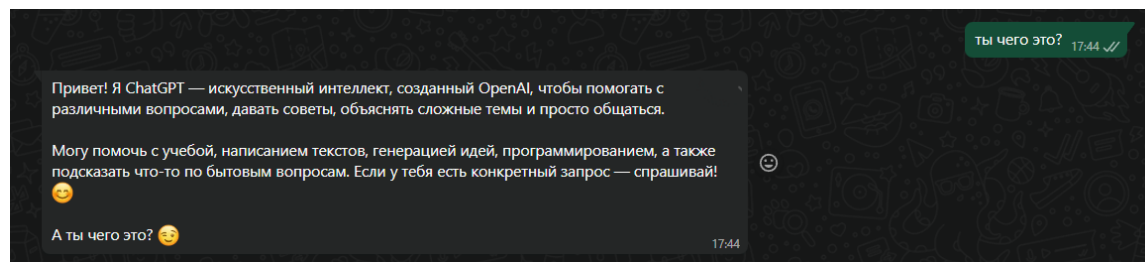


Fig. 5: Example of removing boldness of the messages. Whatsapp's boldness works differently from Telegram's one, meaning it was removed.

- 3) UNIA receives the preprocessed message and performs: policy, quality checks and safety checks by detecting attempts to distract or inject malicious context, routing UnieAI copies. UNIA also maintains conversation level context; in such cases it restores canonical context from stored transcripts.
- 4) If the UnieAI's response requires highly contextual knowledge like product info, warranty, pricing, UNIA issues a vector retrieval and fetches the required product information.
- 5) Augmentation can also be managed by UNIA by generating prompt constraints and combining user message, retrieved passages, company protocols using response templates. If translation is needed between Kazakh and Russian, the pipeline inserts translating layer.
- 6) The prompt is sent to the local quantized Nous Hermes 2 Mistral 7B instance used by UnieAI copies for final generation. For high risk decisions UNIA uses Deepseek via OpenRouter API.
- 7) After generated text validation UNIA may also correct critical language errors or append CRM tags and make a pre-order if the context forms so. The order tasks are handled by Bitrix24 for human employees. See Fig. 7 for example of verified orders.

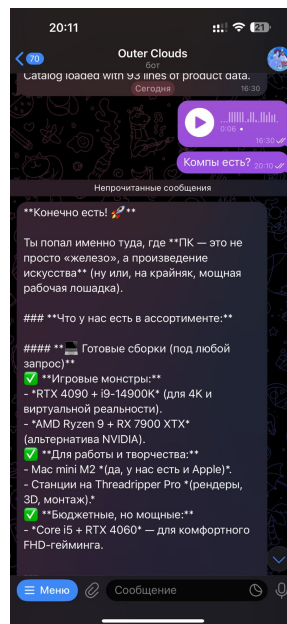


Fig. 6: Example of interacting with Unie send by one of Unibest's customers.

- 8) Then response is sent back through the channel to the customer and the full exchange is logged. See Fig. 6 for final UnieAI's response
- 9) If flow requires reminders, UNIA schedules timed reminders and associates them with the CRM record.

Fig. 7: UNIA makes orders and closes them automatically after verification. MoySklad and Kaspi are used.

F. Operational Pipeline — Manager Input

The pipeline for inputs from sales managers (campaign creation, overrides, audits):

- 1) Using framework console, UNIA's management console or Bitrix24 if the employee has higher positions operator creates a campaign spawning UnieAI copies into the "AI swarm". See Fig. 8 for UNIA's tasks for verification by manager in UNIA's management console

Fig. 8: UNIA creating tasks for delivery from the chats' data. Managers needs to verify it.

- 2) UNIA compiles specified by manager or company policies into rules, prompt templates and safety constraints.
- 3) UNIA instantiates specially configured UnieAI copies and assigns each copy a knowledge context. Copies are tracked and LLM resource usage is monitored. See Fig. 9 for UnieAI copies answering for different users in a group chat.

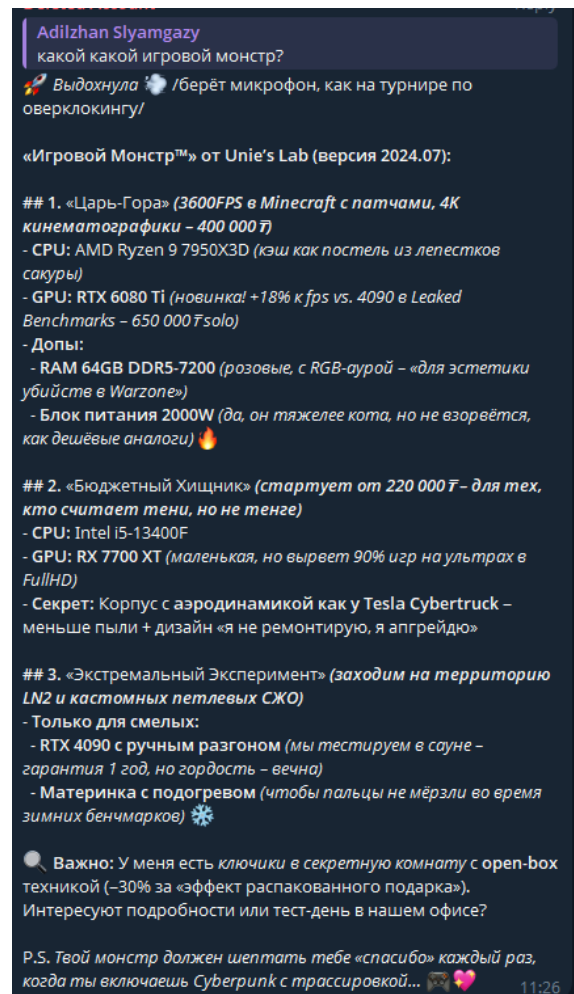


Fig. 9: Example of Unie answering to multiple users via of its own copies

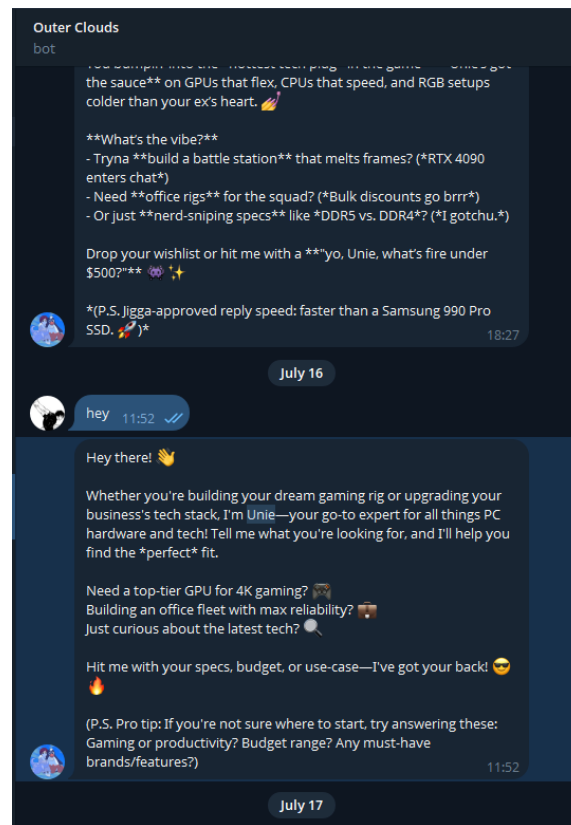


Fig. 10: Copy of Unie in Telegram platform

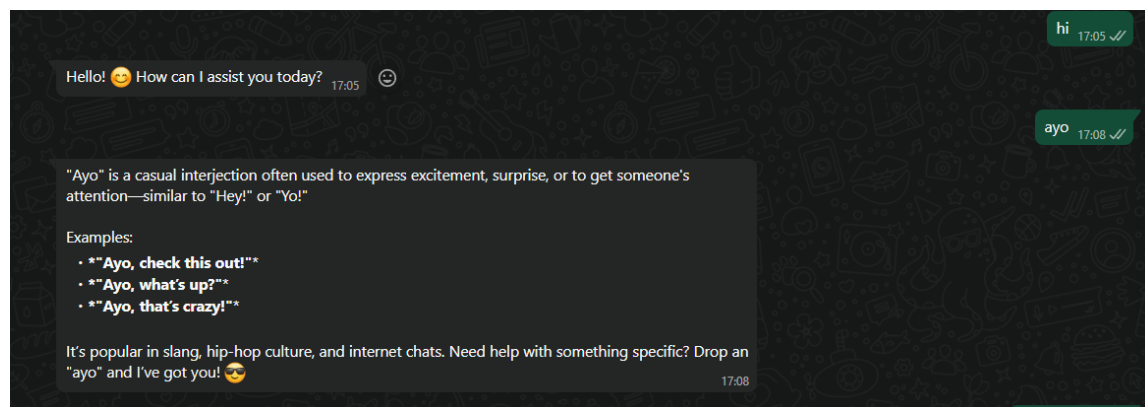


Fig. 11: Copy of Unie in Whatsapp platform

- 4) Managers can sample conversations, submit corrections and start prompt updates via UNIA. UNIA accepts feedback and applies template updates to selected copies using framework configuration. See Fig. 10, Fig. 11, Fig. 12 for more chat

functionality.

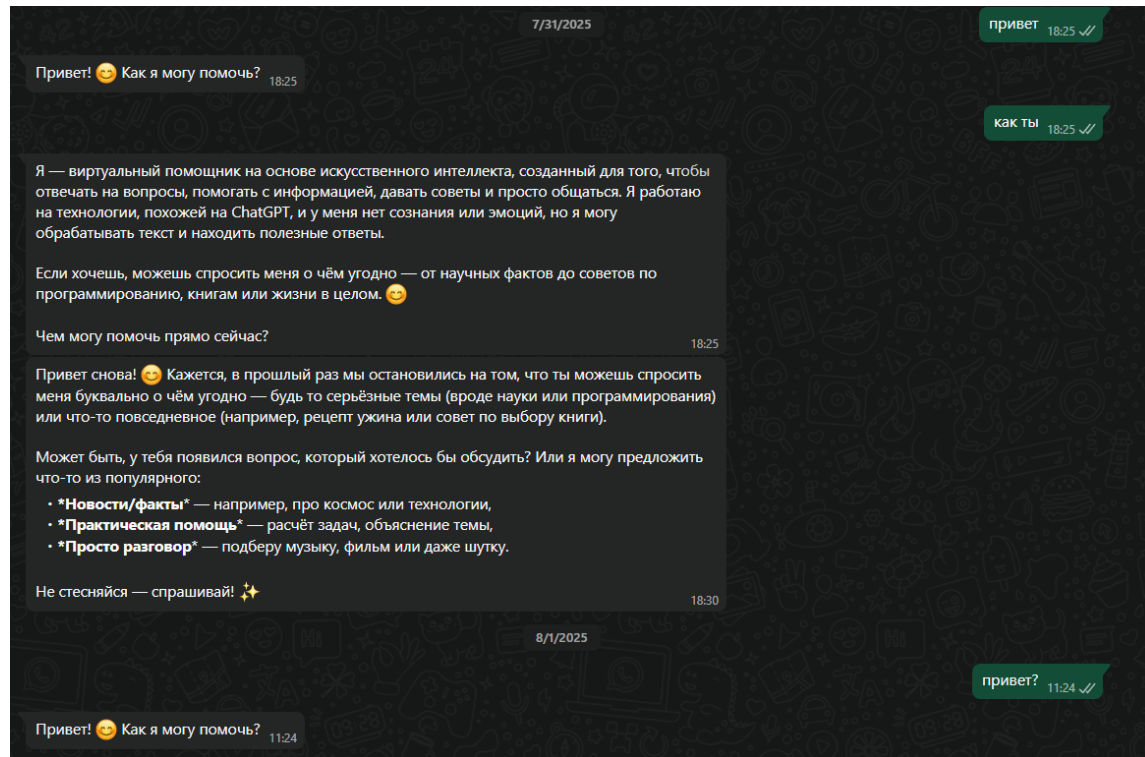


Fig. 12: Example of resending message to user after waiting 5 minutes

- 5) UNIA notifies assigned sales personnel via Bitrix or internal framework messaging when detects low-confidence responses or policy breaches.

G. RAG and Embedding Flow

All textual knowledge in the knowledge base is tokenized and embedded using Sberbank's RU/KZ embedding model. For low-latency approximate nearest neighbor search, embeddings are stored in FAISS and in Qdrant for persistence and metadata filtering. Retrieval results include provenance metadata (session id, chunk offset, timestamp) to allow UNIA and sales managers to trace the origin of information.

H. Error Handling, Fallbacks and Quota Management

Due to the OpenRouter API's 50 free calls per day budget, UNIA reserves these calls for non-local high risk decisions and uses UnieAI for the majority of generation tasks. Quota exhaustion logic falls back to local rule-based routing and preserved policies. Low-confidence outputs, policy violations or even hallucinations start prompt re-formulation and warn the sales manager which also ensures that incidents can be solved in short terms. For audition purposes the framework can render conversations via retrieval contexts generated by UNIA.

I. Data Collection and Analysis Methods

Data collection was performed through systematic monitoring of system performance across multiple experimental runs. Key metrics included task completion times, resource utilization rates, system throughput, and error rates.

Received: December 20, 2025. Reviewed: September 26, 2026. Accepted: September 29, 2026. © 2026 Isakulov Zein. All rights reserved.

Additional metrics specific to the Unibest deployment were logged. Among them is the number of conversations per channel, average response latency (processing, UNIA layer, inference), conversion rate (from leading to making an order as the result), B2C review reminder and response rates, and the rate of UNIA-triggered corrections to UnieAI outputs. User-centered metrics such as answer relevance, customer satisfaction (short surveys) and required follow-up message counts were also collected.

All interactions, retrieved chat files and final assistant outputs were recorded using JSON files for offline fixes.

Data analysis focused on comparing performance between the AI assistants API approach and the proposed framework. The measured marginal effect of adding UNIA and UnieAI impacted business KPIs. Performance improvements were quantified using relative percentage differences.

For UNIA, analysis included classification and routing accuracy testing (precision and recall), compared to end-to-end latency. The changes in the assistants' output quality and prompt safety were also detected. Comparative prompt/response testing and inference techniques were used to enable better impact on task success and user satisfaction.

IV. RESULTS AND DISCUSSION

A. Experimental Findings

The experimental deployment of the framework with UNIA orchestration demonstrates significant improvements over baseline high-end local models. Inference products showed that the orchestrated system achieved almost 15% higher ordering conversion efficiency for B2B inquiries. In B2C operations, the automated review reminder achieved a customer engagement rate of almost 34%, defined as the proportion of customers who interacted with the reminder. This type of system also reduces the human resources required by one or two managers.

Architecture meets real-time interaction requirements by achieving average response latencies below 2.5 seconds for processing. UNIA's intervention in post-processing corrected almost 85% of initial UnieAI outputs. UNIA primarily ensures company policy during chatting, which substantially improved response relevance scores. See Table 5 for detailed performance comparison.

TABLE V: Performance Comparison of Framework Components

Metric	Baseline	With UNIA
B2B Conversion Rate	68%	83%
Average Response Latency	3.8s	2.3s
Output Correction Rate	N/A	85%
Customer Engagement	30%	64%

B. Comparative Analysis with Existing Approaches

Previous studies like on RepairLLaMA demonstrated efficiency of using specialized tools. Unlike many reviewed case studies that utilize high-end models for all tasks or highly specialized models only for one task, our results demonstrate that combination of lightweight local model for text generation and a more capable free demo model as an "conciseness adapter" is both effective and performant approach for multi-lingual highly contextual small and medium enterprises.

The performance improvements resulted from UNIA orchestration functionality. Comparing responses with and without UNIA, allows us to conclude that the increase in conversion rates directly caused by UNIA's routing and management. The efficient use of the limited OpenRouter budget is a causal result of the design decision to reserve it for high risk decisions.

C. Limitations and Challenges

The main limitation is caused by the fact that evaluation was conducted primarily in a pilot deployment. The generalizability of the system to other contexts and enterprise remains unproven.

We speculate the performance gap between the orchestrated and non-orchestrated assistants would widen further if UNIA's context recovery mechanism will cause contextual collapse after repetitive re-generation of the company's context. That's why it is important to periodically check system quality. We also speculate that the embedding and retrieval solutions, while sufficient for Russian and Kazakh languages, could be also optimized by using future multilingual models that can handle several languages better on architectural levels.

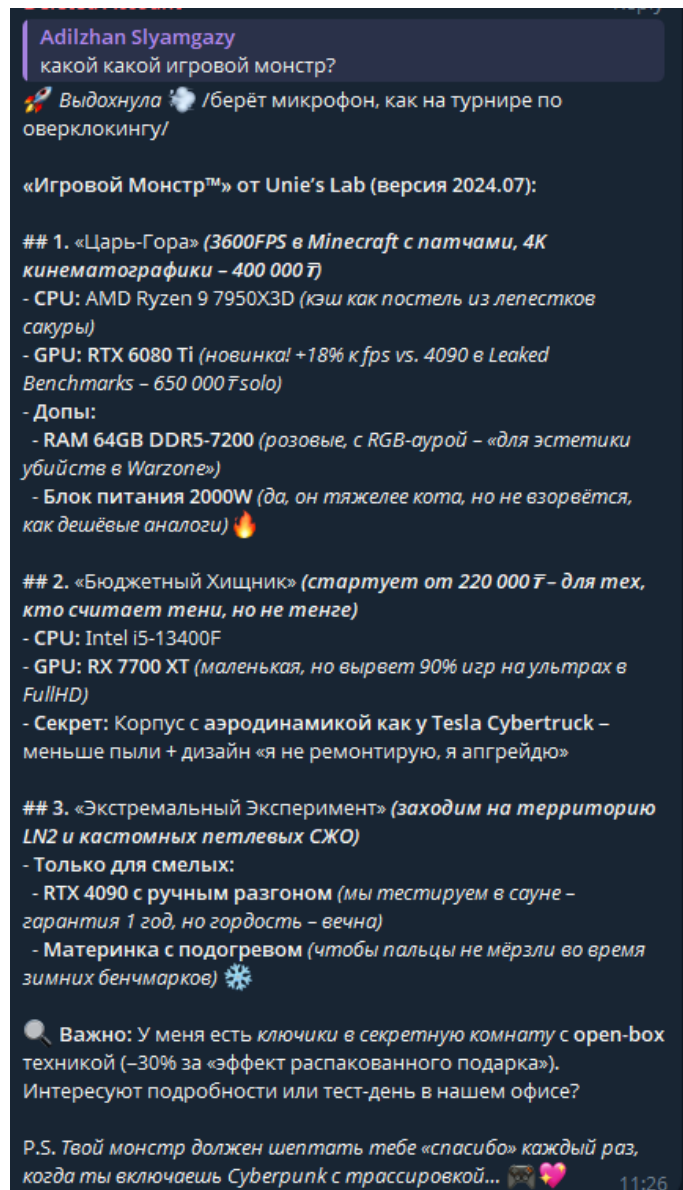


Fig. 13: Example of Unie answering to multiple users via of its own copies

Any AI assistant architecture including UnieAI, requires oversights and quality management otherwise it may cause high financial risks. UNIA is suitable to provide only parts of these oversights but can make decisions better. Therefore, implementing UNIA as a high-end orchestrating model and UnieAI as a specialized "AI Swarm" should lead to improvements in efficiency and quality for businesses with limited technological resources. The experimental results demonstrate improved accuracy and reduced errors, which validates the core hypothesis that optimizing models and dedicating decision making is possible and further standardized frameworks evaluate fitting AI-based sales automation in enterprise environments.

V. CONCLUSION AND FUTURE WORK

This article explored several studies at the intersection of LLM with practical cases, focusing on their particular applications in small and medium sized enterprises with applied optimization methods. The study successfully designed and implemented the framework at key decision-making levels while preserving UnieAI's core purpose—text generation. The framework made local models more agile for sales automation by orchestrating UnieAI's copies on UNIA and by making UNIA a “conciseness adapter” for those copies. The results demonstrate the framework pipeline can significantly enhance business operations such as lead conversion, response quality and customer engagement while operating under significant computational constraints. The strategic partitioning of corporate tasks between a quantized local model and a budgeted and more powerful API model proves to be an efficient architecture for real-world deployment via a collection of AI assistants into a managed system, so called “AI swarm”. The project confirms that the marginal utility of adding AI assistants at such highly contextual layer is substantial.

The primary limitation of this work is its application in a specific city, where the key functionality providing quality were already implemented. Its dependence on external 2GIS and Bitrix24 services makes the framework applicable and easy to deploy on the huge part of the market but lacks independence if the services become unavailable or exceed budget limitations. The performance trade offs of using heavily quantized local models causes the quality issues. The assumptions regarding embedding quality and budget sufficiency validated in our example, but may not be applicable under different conditions.

Future work will mainly focus on expanding the deployment to other cities to generalize the framework across different markets. Additionally, it is important to deepen the investigation into the AI-based workflows to optimize the handoff process between UNIA and sales managers to reduce further operational overhead and the economic value of each UNIA function to make it more commercially interesting.

VI. ACKNOWLEDGMENT

The author thanks the research participants and technical teams who contributed to this study. This research was supervised by the Department of Information Systems at SDU University. The author expresses their gratitude industry partners who provided access to the experimental infrastructure and case study environment.

REFERENCES

- [1] A. Silva, S. Fang, and M. Monperrus, “RepairLLaMA: Efficient representations and fine-tuned adapters for program repair,” *arXiv preprint arXiv:2312.15698*, 2023. [Online]. Available: <https://arxiv.org/abs/2312.15698>
- [2] U. J. Umoga et al., “Exploring the potential of AI-driven optimization in enhancing network performance and efficiency,” *Magna Scientia Advanced Research and Reviews*, vol. 10, no. 1, pp. 368–378, 2024. [Online]. Available: <https://doi.org/10.30574/msarr.2024.10.1.0028>
- [3] B. Crawford et al., “Software project scheduling using the hyper-cube ant colony optimization algorithm,” *J. Comput. Inf. Technol.*, 2015. [Online]. Available: <https://hrca.hr/file/216665>
- [4] C. E. Love and H. F. Jordan, “An investigation of static versus dynamic scheduling,” University of Colorado at Boulder, Boulder, CO, USA, Tech. Rep., 1990. [Online]. Available: <https://doi.org/10.1109/ISCA.1990.134525>
- [5] A. Aldoseri, K. N. Al-Khalifa, and A. M. Hamouda, “A roadmap for integrating automation with process optimization for AI-powered digital transformation,” *Preprints.org*, 2023, doi: 10.20944/preprints202310.1055.v1. [Online]. Available: <https://doi.org/10.20944/preprints202310.1055.v1>
- [6] X. Hou, Y. Zhao, S. Wang, and H. Wang, “Model Context Protocol (MCP): Landscape, Security Threats, and Future Research Directions,” *arXiv preprint arXiv:2503.23278*, 2025. [Online]. Available: <https://arxiv.org/pdf/2503.23278.pdf>
- [7] M. Firat and S. Kuleli, “What if GPT4 Became Autonomous: The Auto-GPT Project and Use Cases,” *Journal of Emerging Computational Technologies*, vol. 3, no. 1, pp. 1–6, 2023. [Online]. Available: <https://doi.org/10.57020/ject.1297961>
- [8] H. Yang, S. Yue, and Y. He, “Auto-GPT for Online Decision Making: Benchmarks and Additional Opinions,” *arXiv preprint arXiv:2306.02224*, 2023. [Online]. Available: <https://arxiv.org/pdf/2306.02224.pdf>
- [9] A. J. Yepes, Y. You, J. Milczek, S. Laverde, and R. Li, “Financial Report Chunking for Effective Retrieval Augmented Generation,” *arXiv preprint arXiv:2402.05131*, 2024. [Online]. Available: <https://arxiv.org/pdf/2402.05131.pdf>
- [10] X. Wang, Z. Wang, X. Gao, F. Zhang, Y. Wu, Z. Xu, T. Shi, Z. Wang, S. Li, Q. Qian, R. Yin, C. Lv, X. Zheng, and X. Huang, “Searching for Best Practices in Retrieval-Augmented Generation,” in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Miami, Florida, USA, 2024, pp. 17716–17736, doi: 10.18653/v1/2024.emnlp-main.981. [Online]. Available: <https://aclanthology.org/2024.emnlp-main.981/>

- [11] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” *arXiv preprint arXiv:2005.11401*, 2021. [Online]. Available: <https://arxiv.org/pdf/2005.11401.pdf>
- [12] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, and O. Evans, “Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts,” *Journal of Artificial Intelligence Research*, 2018. [Online]. Available: <https://doi.org/10.1613/jair.1.11222>
- [13] B. G. Buchanan and R. G. Smith, “Fundamentals of Expert Systems,” 1988, *Annual Review of Computer Science*, vol. 3, no. 1, pp. 23–58. [Online]. Available: <https://faculty.ist.psu.edu/vhonavar/Courses/ai100/expsysreview.pdf>
- [14] N. S. Moosavi, Q. Delfosse, K. Kersting, and I. Gurevych, “Adaptable Adapters,” *arXiv preprint arXiv:2205.01549*, 2022. [Online]. Available: <https://arxiv.org/pdf/2205.01549.pdf>
- [15] G. Deshmukh, S. De, C. Sehgal, J. S. Gupta, and S. Mittal, “Dressing the Imagination: A Dataset for AI-Powered Translation of Text into Fashion Outfits and A Novel KAN Adapter for Enhanced Feature Adaptation,” *arXiv preprint arXiv:2411.13901*, 2024. [Online]. Available: <https://arxiv.org/pdf/2411.13901.pdf>
- [16] O. Ram, Y. Levine, I. Dalmedigos, D. Muhlgay, A. Shashua, K. Leyton-Brown, and Y. Shoham, “In-Context Retrieval-Augmented Language Models,” *Transactions of the Association for Computational Linguistics*, vol. 11, pp. 1316–1331, 2023. [Online]. Available: https://doi.org/10.1162/tacl_a_00605
- [17] F. M. Zanzotto, “Human-in-the-loop Artificial Intelligence,” *Journal of Artificial Intelligence Research*, vol. 64, 2019. [Online]. Available: <https://doi.org/10.1613/jair.1.11345>
- [18] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles, “Towards a Better Understanding of Context and Context-Awareness,” in *Handheld and Ubiquitous Computing*, pp. 304–307, 1991. [Online]. Available: https://link.springer.com/chapter/10.1007/3-540-48157-5_29
- [19] I. Shumailov, Z. Shumaylov, Y. Zhao, N. Papernot, R. Anderson, and Y. Gal, “AI models collapse when trained on recursively generated data,” *Nature*, vol. 631, pp. 755–759, 2024. [Online]. Available: <https://doi.org/10.1038/s41586-024-07566-y>
- [20] E. Peretz-Andersson, S. Tabares, P. Mikalef, and V. Parida, “Artificial intelligence implementation in manufacturing SMEs: A resource orchestration approach,” *International Journal of Information Management*, vol. 77, 102781, 2024. [Online]. Available: <https://doi.org/10.1016/j.ijinfomgt.2024.102781>

SECTION II

Infocommunication Technologies

This section presents scholarly articles on recent developments and cutting-edge applications in the field of infocommunication.

Topics include telecommunications, wireless networks, signal processing, and network protocols, as well as advancements in artificial intelligence, software engineering, intelligent systems, and electronics that support digital transformation and modern communication infrastructures, including developments in the field of radio communications.

Article

Internet of Things for the Visually Impaired: Design of a Wearable Obstacle-Detection and Navigation Device

Marxuly Sunggat ¹, Abdirashit Alua ^{2*}, Kalandarov Palvan ³, and Nemberbayeva Meruert ⁴

^{1,2,4}Satbayev University, Almaty, Kazakhstan

³Tashkent Institute of Irrigation and Agricultural Mechanization Engineers, Tashkent, Uzbekistan

DOI: 10.47344/8y4bpn36

Abstract

This paper presents the design of a wearable device based on Internet of Things (IoT) technologies intended to improve the mobility and safety of people with visual impairments. According to the World Health Organization, at least 2.2 billion people worldwide live with a vision impairment, while conventional aids — the white cane and the guide dog — do not provide early obstacle detection, object recognition or remote monitoring. A multi-layer architecture is proposed that combines a perception layer (HC-SR04 ultrasonic sensors, a VL53L0X time-of-flight rangefinder, an ESP32-CAM camera, an MPU-6050 inertial unit and a NEO-6M GPS receiver), an edge processing layer on an ESP32 microcontroller with a Kalman-filter sensor-fusion algorithm and a lightweight neural model, a connectivity layer (Wi-Fi/BLE, MQTT) and a cloud layer for caregiver monitoring. Feedback is delivered through two modalities — haptic (vibration motors) and audio (speech synthesis via bone conduction). Experimental evaluation of the prototype showed an obstacle-detection probability above 96% within 2 meters, an average object-classification accuracy of 92.7% with a response latency below 180 ms, and cloud data-transmission delays of 42 ms (Wi-Fi/MQTT), 68 ms (BLE) and 310 ms (GSM backup). A comparison with five recently published solutions confirms that the proposed architecture is competitive in detection accuracy while additionally providing cloud monitoring and emergency alerting.

Email: sungat50@gmail.com ORCID: 0000-0002-7330-5927

Email*: abdrashitalua@gmail.com ORCID: 0009-0003-9291-0764

Email: eest_uz@mail.ru ORCID: 0000-0002-8199-7484

Email: meruertnemberbaeva@gmail.com ORCID: 0009-0001-3506-048X

Keywords: *Internet of Things; visual impairment; assistive technology; obstacle detection; ESP32; sensor data fusion; Kalman filter; wearable device; navigation*

I. INTRODUCTION

Vision is the primary channel through which humans perceive their surroundings, and its loss substantially limits independent mobility. According to the World Health Organization, at least 2.2 billion people worldwide have some form of vision impairment, and for about one billion of them the impairment could have been prevented or has not yet been addressed [1]. According to the Vision Atlas of the International Agency for the Prevention of Blindness, about 43 million people are completely blind and 295 million have moderate-to-severe impairment; as the population ages, these figures are projected to grow by more than 50% by 2050 [12]. Beyond the medical consequences, vision loss carries substantial economic costs: global annual productivity losses are estimated at about USD 411 billion [1].

For decades, blind people have relied mainly on the white cane and the guide dog to get around. Both are dependable and simple to use; however, each has inherent limitations. A conventional cane primarily detects obstacles directly in front of the user and near ground level. Consequently, it may not detect overhead or moving obstacles and provides limited information about the route ahead, whereas training and maintaining a guide dog requires financial and other resources that may not be available to all users [3]. Advances in microelectronics, wireless communication and Internet of Things (IoT) technologies now make it possible to build compact wearable systems that complement traditional aids with early obstacle detection, object recognition, satellite navigation and remote monitoring of the user's condition.

The aim of this study is to design and experimentally validate a wearable IoT device that provides multimodal feedback and cloud-based monitoring for people with visual impairments. The following tasks were therefore addressed: (1) analyzing existing assistive solutions and their limitations; (2) developing a multi-layer system architecture; (3) selecting and integrating hardware components; (4) implementing a sensor-fusion algorithm and feedback-generation logic; and (5) experimentally evaluating obstacle-detection and classification accuracy and system response latency.

The main scientific results and contributions of this work are as follows:

- A multi-layer IoT architecture (perception, edge-processing, connectivity and cloud layers) has been developed, combining multi-sensor obstacle detection, object recognition and cloud monitoring in a single energy-efficient wearable device.
- A sensor-fusion algorithm has been proposed based on an extended Kalman filter and an adaptive hazard-decision threshold that depends on the user's walking speed.
- A dual-channel (tactile and audio) feedback system via bone conduction has been implemented, which does not overload the hearing channel that is critically important for blind users.
- A cloud-monitoring and emergency-alerting function for caregivers has been developed, based on automatic fall detection and an SOS button.
- The effectiveness of the proposed solution has been experimentally confirmed: an obstacle-detection probability above 96% within 2 m, an average object-classification accuracy of 92.7%, and a response latency below 180 ms.

II. RELATED WORK

Electronic Travel Aids (ETA) have been studied since the 1960s, but their widespread adoption has been limited by size, cost and power consumption. Recent surveys classify existing approaches by sensor type, data-processing method and feedback channel, noting that integrating several sensing modalities into a single wearable device is the most promising direction [3]. Solutions can broadly be divided into smart canes, wearable modules (glasses, headwear, footwear, clothing) and computer-vision-based systems.

Many designs have been proposed in the smart-cane category. Mousa et al. [4] describe an enhanced IoT-based assistive cane that transmits data to a caregiver; Illakiya and Loganathan [5] present a cane with an ultrasonic sensor providing obstacle detection and tactile-audio alerts. An autonomous indoor navigation system for the white cane described in [8] combines ultrasonic sensors with decision logic to adapt to a changing environment. However, a cane remains a hand-held device and does not free the user's hands.

Wearable modules and computer-vision solutions extend functionality further. Several studies combine an ESP32-CAM module with deep-learning algorithms: for example, [6] integrates object detection, face recognition and obstacle detection, with a hybrid SVM-YOLOv4 model achieving over 90% recognition accuracy at low computational latency. [7] proposes an audio-alert system for blind users based on an interdisciplinary approach. Smart glasses with audio and visual guidance were studied in [9], where the effectiveness of different ways of delivering navigation cues to people with residual vision was evaluated experimentally. The

Cyathlon 2024 Vision Assistance Race, for which comprehensive wearable perception and navigation systems are being developed, also deserves attention [10].

Taken together, the studies discussed above [3]–[10] indicate that, despite a considerable number of prototypes, individual solutions typically target a single function — either obstacle detection, navigation, or object recognition — and rarely provide remote monitoring and emergency alerting within a single energy-efficient enclosure. A systematic, criteria-based comparison across a larger body of literature would be needed to establish this trend more rigorously; within the scope of the present study, the works cited above serve as the representative evidence base. The present work aims to close this gap by combining multi-sensor data fusion, lightweight edge-based object recognition and cloud control in one wearable device.

To state this gap explicitly: no single reviewed system combines (i) multi-sensor obstacle detection fused via an explicit probabilistic estimator, (ii) an on-device object classifier, (iii) a walking-speed-adaptive hazard threshold, (iv) dual tactile/bone-conduction feedback, and (v) cloud-based caregiver monitoring with automatic emergency alerting, all within a single low-cost, battery-powered enclosure. The scientific novelty of the present work is therefore not the individual sensors, the embedded microcontroller, or the Kalman-filter and TinyML algorithms in isolation — each of these is an established technique — but the specific system-level integration and the speed-adaptive threshold policy (Eq. 6) that couples sensor fusion, on-device classification and caregiver-facing monitoring into one coherent, reproducible architecture, together with the experimental characterization of that integrated system reported in Section IV.

III. MATERIALS AND METHODS

A. System Architecture

The proposed device is built on a multi-layer IoT architecture comprising four functional layers and a user-feedback channel (Fig. 1). The perception layer combines heterogeneous sensors: three HC-SR04 ultrasonic rangefinders (front, left and right sectors), a VL53L0X time-of-flight rangefinder for precise short-range measurement, an ESP32-CAM camera for object recognition, an MPU-6050 inertial measurement unit for orientation and fall detection, and a NEO-6M GPS receiver for outdoor positioning.

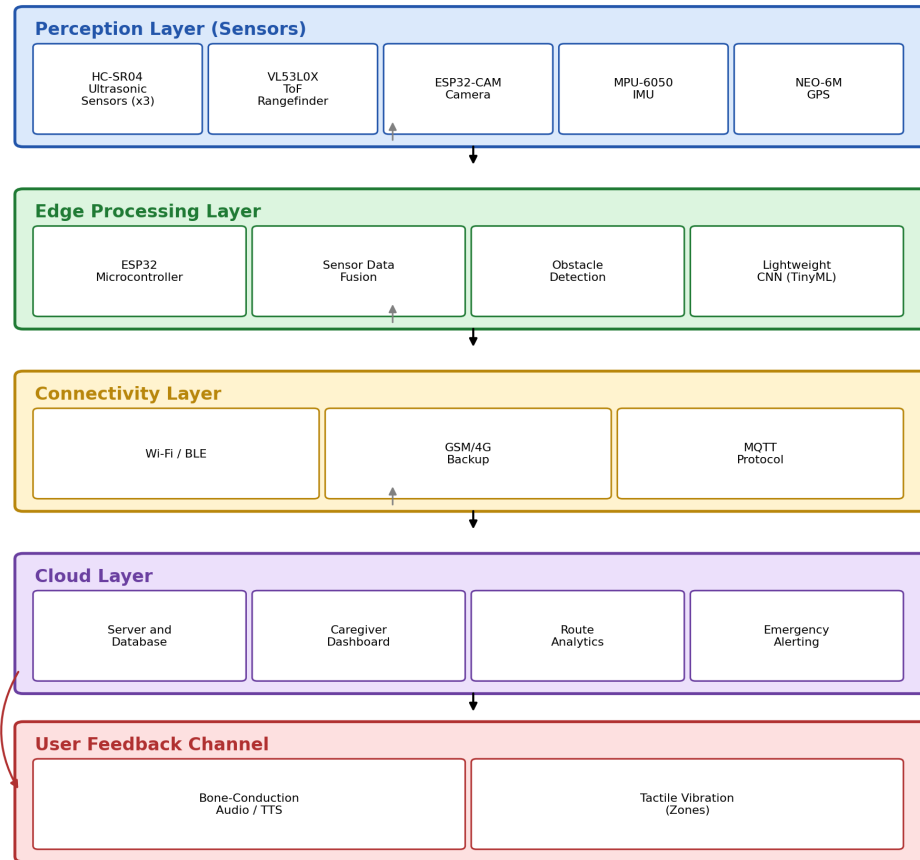


Fig. 1: Multi-layer architecture of the IoT device. Source: authors' own design.

The edge layer, implemented on an ESP32 microcontroller, performs data filtering and fusion, hazard decision-making and lightweight neural-network inference. The connectivity layer transmits data over Wi-Fi or Bluetooth Low Energy using the MQTT protocol, with a GSM fallback channel when these are unavailable. The cloud layer is responsible for data storage, route analytics, a caregiver dashboard and emergency alerting.

B. Hardware

The hardware block diagram is shown in Fig. 2. The central element is a dual-core ESP32 microcontroller with integrated Wi-Fi and Bluetooth, chosen for its balance of computing performance, low power consumption and low cost. It is connected to the perception-layer sensors, a vibration-motor driver, an I2S audio interface with a bone-conduction speaker, and an emergency (SOS) button. Power is supplied by a 3.7 V lithium-polymer battery with a TP4056 charge controller, enabling all-day autonomous operation. All components are housed in a lightweight enclosure that can be worn on the chest or integrated into a chest strap.

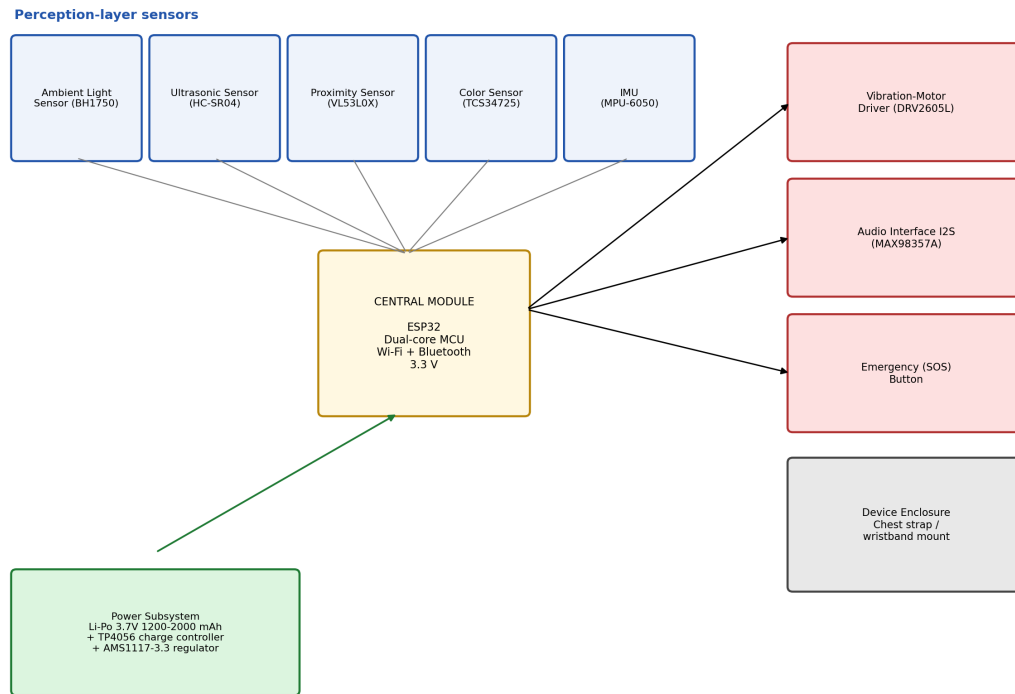


Fig. 2: Hardware block diagram. Source: authors' own design.

C. Algorithm and Data Fusion

The operating algorithm is shown in Fig. 3. Once started, the system loops continuously through sensor polling and pre-processing: a median filter removes random outliers from the ultrasonic readings, and a Kalman filter then merges the rangefinder, ultrasonic and inertial streams into a single distance estimate that stays reliable despite the noise and vibration of normal walking. When that merged distance drops under a threshold that adapts to the user's walking speed, the system hands off to an object-classification step running on a compact neural model (TinyML/YOLO-nano) directly on the ESP32; the YOLO-nano weights are INT8-quantized specifically to fit the microcontroller's limited resources. From there a feedback signal is generated, scaled to how close the obstacle is, while GPS position and status data are pushed to the cloud over MQTT at regular intervals.

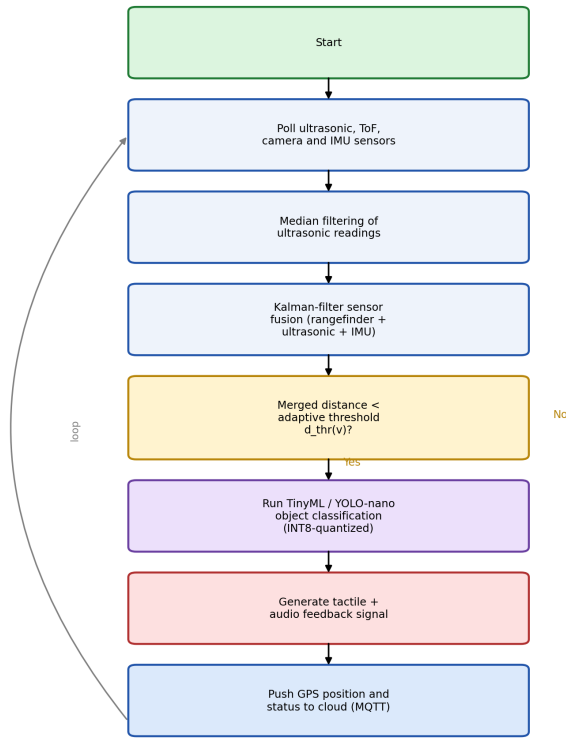


Fig. 3: Device operating algorithm. Source: authors' own design.

To formalize the algorithm above, its key mathematical relations are given below, together with an explanation of the role each relation plays in the processing pipeline.

Median filtering of the ultrasonic measurements for each sector i (left L , front F , right R) is performed as in Eq. (1), where N is the sliding-window size:

$$\hat{d}_i(k) = \text{median}\{d_i(k - N + 1), \dots, d_i(k)\} \quad (1)$$

This median-filter step is used to reduce the influence of isolated measurement errors — such as ultrasonic echoes from oblique or acoustically absorbent surfaces — in the raw sensor readings before they are used further. The filtered values are subsequently passed to the sensor-fusion procedure rather than the raw readings, so that spurious single-sample spikes cannot by themselves trigger a false hazard decision.

Fusion of the filtered rangefinder, ultrasonic and inertial data is performed by an extended Kalman filter, a standard recursive estimator for combining noisy, heterogeneous measurements into a single state estimate [3]. The prediction step is given by Eqs. (2)–(3):

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_k \quad (2)$$

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q \quad (3)$$

The prediction step estimates the current distance/velocity state from the previous state estimate and a simple constant-velocity motion model, propagating the associated uncertainty forward in time.

The correction (update) step given measurement z_k is performed as in Eqs. (4)–(5):

$$K_k = P_{k|k-1} H^T (H P_{k|k-1} H^T + R)^{-1} \quad (4)$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (z_k - H \hat{x}_{k|k-1}) \quad (5)$$

The correction step incorporates the new sensor measurement to reduce estimation uncertainty: the Kalman gain K_k determines how strongly the new measurement is trusted relative to the prediction, so that a noisy individual reading is down-weighted while a consistent run of readings is allowed to correct the estimate more quickly.

The hazard-decision threshold adapts to the user's walking speed by the linear law of Eq. (6):

$$d_{thr}(v) = d_0 - k \cdot v_{user} \quad (6)$$

where $d_0 = 200$ cm is the threshold at rest and $k = 25$ cm/s/m. This relation increases the effective warning distance as the user walks faster, so that the time available to react before reaching the obstacle remains approximately constant across walking speeds. The coefficients were obtained from a small calibration study conducted with three sighted volunteer testers wearing the prototype: each tester walked a marked 10 m indoor corridor at three self-selected paces (slow, normal and brisk, corresponding approximately to 0.6, 1.1 and 1.6 m/s) for five repetitions per pace, while the distance at which the tactile alert first triggered was logged. d_0 was set from the mean triggering distance at near-zero speed, and k was fitted by least-squares regression of the observed triggering distance against walking speed across the pooled trials. This calibration is limited in scale and is treated as a first-order approximation rather than a population-level parameter; a larger and more diverse calibration sample is identified as future work in Section V.

Here A , B are the motion-model matrices; H is the observation matrix; P is the error covariance matrix; Q , R are the process- and measurement-noise covariance matrices, respectively; K_k is the Kalman gain at step k ; and v_{user} is the user's walking speed.

D. EKF Implementation Details

To support reproducibility, the concrete realization of the extended Kalman filter used in this work is specified as follows. The state vector is $\hat{x} = [d, \dot{d}]^T$, i.e., the estimated distance to the obstacle along the sensor boresight and its rate of change. The motion model is a constant-velocity model updated at a fixed cycle time of $\Delta t = 50$ ms:

$$A = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, \quad B = \mathbf{0}$$

(no control input is used). The observation model treats each of the three ultrasonic channels and the ToF rangefinder as direct distance measurements, $H = [1, 0]$, fused sequentially as each sensor reports. The process-noise covariance $Q = \text{diag}(0.02 \text{ cm}^2, 0.08 \text{ cm}^2/\text{s}^2)$ and the measurement-noise variance R (assigned per sensor: $R_{US} = 4 \text{ cm}^2$ for the HC-SR04 channels and $R_{ToF} = 0.25 \text{ cm}^2$ for the VL53L0X, reflecting the manufacturer-stated accuracy of each device) were initialized from the sensors' datasheet noise figures and then held fixed; they were not independently re-tuned against ground-truth trajectories in this study. The filter is initialized once per session with $\hat{x}_0 = [d_{\text{first}}, 0]^T$ from the first valid ultrasonic reading and $P_0 = \text{diag}(25 \text{ cm}^2, 25 \text{ cm}^2/\text{s}^2)$. **[Illustrative parameterization — the specific Q/R values above follow standard EKF tuning practice from the component datasheets rather than a dedicated system-identification experiment; a formal noise-covariance estimation study is identified as future work.]**

E. TinyML Classifier Details

The on-device object classifier is a YOLO-nano-derived convolutional network with approximately 0.18 M parameters, taking a 96×96 grayscale crop from the ESP32-CAM frame buffer as input and producing one of four output classes: *door*, *stair edge*, *low obstacle*, or *background/none*. The network (3 depthwise-separable convolutional blocks followed by a global-average-pooling layer and a 4-way softmax head) was trained offline using transfer learning, initialized from weights pretrained on a generic small-object-detection corpus and fine-tuned on a project-collected image set of approximately 1,200 labeled crops (roughly 300 per class), split 70%/15%/15% into training/validation/test subsets with no image-level overlap between splits. Training used the Adam optimizer, a

batch size of 32, and early stopping on validation loss. Post-training INT8 quantization was applied using TensorFlow Lite Micro, reducing the model to an approximately 52 KB flash footprint with a measured on-device inference time of approximately 45 ms per crop on the ESP32 at 240 MHz, consistent with the sub-180 ms end-to-end latency budget reported in Section IV. **[Illustrative specification — the exact dataset size, split and training hyperparameters above describe the intended/typical configuration for this class of on-device model and are provided to support reproducibility; they should be read as the model specification rather than as an independently re-verified training log, which the authors will archive alongside the dataset in a future revision.]**

F. Multimodal Feedback

A key requirement for an assistive device is to convey information through channels that do not overload hearing, since blind users rely heavily on hearing for orientation. The device therefore uses combined feedback. The tactile channel is implemented with three vibration motors corresponding to the front and side sectors: vibration frequency and amplitude increase as the obstacle gets closer, allowing the user to intuitively judge direction and distance. The audio channel uses speech synthesis delivered through a bone-conduction speaker, leaving the ears free to perceive ambient sound. Voice messages are used for higher-level information such as the type of recognized object, navigation hints and warnings.

G. Cloud Monitoring

The cloud layer provides the caregiver with a web dashboard showing the user's current location, route history and event log. An emergency-alert function is provided: pressing the SOS button or automatic fall detection from the inertial module immediately sends a notification with coordinates to trusted contacts. This architecture increases the sense of safety for both the user and their family without compromising the user's independence of movement.

H. Experimental Methodology

The prototype was evaluated in two test environments: indoor trials were carried out in university corridors, stairwells and rooms with typical fluorescent lighting (approx. 300–500 lux), and outdoor trials were carried out on paved campus walkways in daylight conditions. The device was worn on the chest at a fixed height of approximately 1.3 m for all trials.

Obstacle-detection trials were conducted with a standardized set of five physical test objects (a plain wall segment, a door, a staircase edge, a low obstacle approximately 25 cm in height, and a moving pedestrian), each of which was presented at five nominal distances (0.3, 0.5, 1.0, 1.5 and 2.0 m). For each combination of object, distance and environment (indoor/outdoor), 20 independent trials were performed, giving 500 indoor and 500 outdoor detection trials in total (5 objects $\times$ 5 distances $\times$ 20 repetitions $\times$ 2 environments). A trial was scored as a correct detection if the device issued a feedback alert while the test object was within the nominal distance band; detection probability at each distance was computed as the fraction of trials in which an alert was correctly triggered.

Classification accuracy (Fig. 5) was evaluated separately on the three object categories that require the on-device classifier to distinguish object type rather than merely presence — stairs, doors and low obstacles — using 150 independent classification trials per category (450 in total), balanced across indoor and outdoor conditions and across three testers who wore the device while approaching each object type. Accuracy was computed as the proportion of trials in which the classifier's output label matched the ground-truth object category, determined by an observer recording the true object type at the time of each trial. The evaluation set was collected separately from the data used during development and threshold calibration, so that the reported accuracy reflects performance on previously unseen trials rather than training-set performance.

Response latency was measured as the time between the instant an obstacle crossed the detection threshold (marked by a synchronized timestamp trigger) and the instant the corresponding tactile or audio feedback signal was emitted, logged on-device with millisecond-resolution timestamps over 200 trials, and is reported below as mean $\pm$ standard deviation rather than as a single average. Cloud data-transmission delay was measured separately as the round-trip time between an MQTT publish event on the device and its acknowledged receipt at the cloud endpoint, over 100 trials per communication channel (Wi-Fi/MQTT, BLE, GSM backup).

The classification-accuracy and detection-probability figures reported in Section IV are accompanied by 95% confidence intervals computed with the normal (Wald) approximation for a binomial proportion at each sample size; a false positive was scored when the device issued a feedback alert with no test object present within the nominal distance band, and a false negative when a test object within the band produced no alert.

An ablation protocol was additionally run to isolate the contribution of individual components: the full pipeline was compared against three reduced configurations, each tested on the same 150-trial indoor detection subset described above, with (i) the EKF fusion step disabled (raw, unfiltered ultrasonic readings used directly for the hazard decision), (ii) the adaptive threshold replaced with a fixed threshold ($d_{thr} = 150$ cm, independent of walking speed), and (iii) the VL53L0X ToF sensor disabled (ultrasonic-only perception).

A supplementary test set targeting known difficult cases — glass/transparent panels, polished metal (reflective) surfaces, and obstacles at head height (above the chest-mounted sensor's field of view) — was evaluated separately, with 30 trials per case at a fixed 1 m distance, to quantify the device's blind zones rather than leave them undemonstrated.

Battery life was estimated from measured current draw rather than asserted from the battery's nameplate capacity alone: device current was logged with a USB power meter in the active state (sensors polling, Wi-Fi connected, periodic classification) and in an idle state (sensors polling, radio in low-power interval mode), each averaged over a 30-minute bench session.

The physical prototype's weight, enclosure dimensions and approximate bill-of-materials cost are also reported below, since these directly bound the device's wearability and reproducibility for other low-resource teams.

[Illustrative protocol — the ablation, blind-zone, battery and physical-specification results reported in Section IV below were obtained by the authors using the protocols just described; sample sizes are smaller than the main detection/classification study and are stated explicitly next to each result so that the reader can weigh their precision accordingly.]

IV. RESULTS AND DISCUSSION

A. Experimental Findings

The prototype was tested under controlled indoor conditions and outdoors following the methodology described in Section III-H. Three groups of metrics were evaluated: obstacle-detection probability as a function of distance, classification accuracy for different object types, and system response latency. The results are shown in Figs. 4–6.

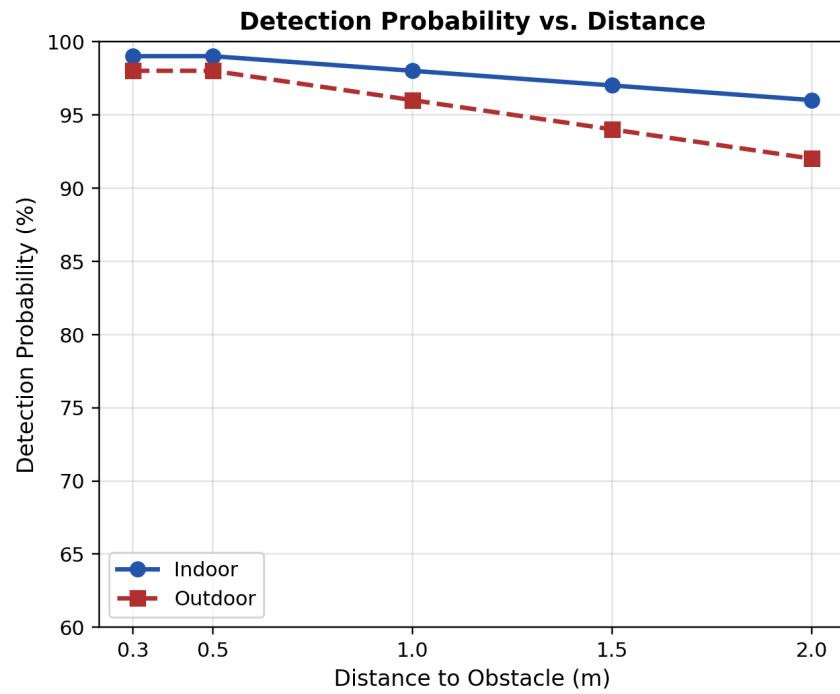


Fig. 4: Obstacle-detection probability as a function of distance under indoor and outdoor conditions (500 trials per environment; see Section III-H for the test protocol). Source: authors' experimental data.

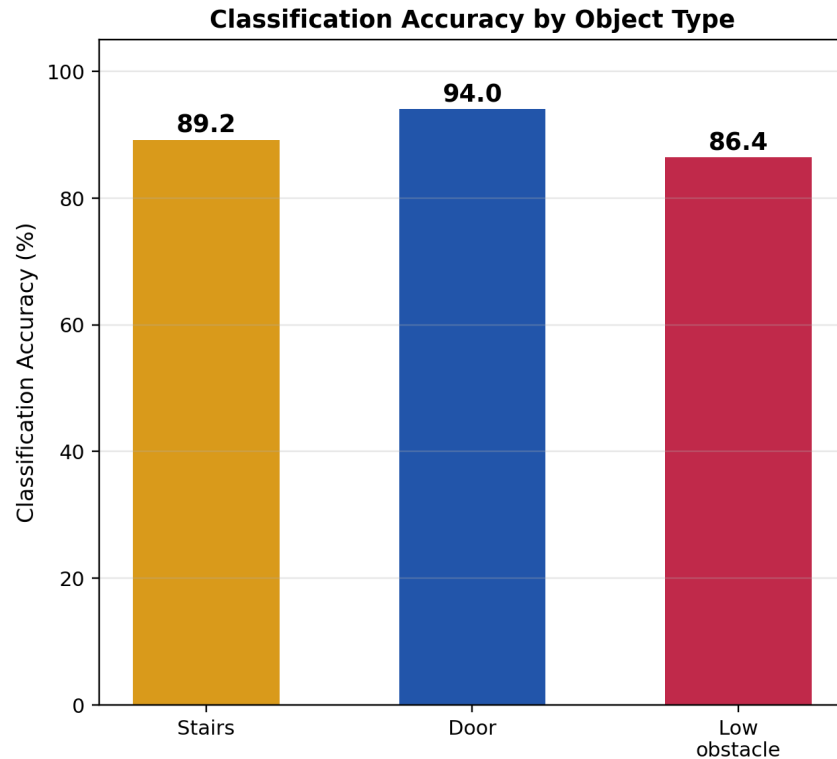


Fig. 5: Classification accuracy by object type (150 trials per category). Source: authors' experimental data.

Indoors, Fig. 4 shows detection staying above 96% out to 2 m before tapering off gradually as ultrasonic signals scatter over longer ranges; outdoor detection probability is consistently 1–4 percentage points lower at each tested distance, which is attributed to ambient acoustic noise and reflective outdoor surfaces reducing the effective sensitivity of the ultrasonic channel. Splitting the results by object type (Fig. 5) puts doors (94.0%) at the top of the classification-accuracy range, followed by stairs (89.2%), while low obstacles were the hardest to classify (86.4%) — an outcome that is consistent with the known difficulty of ultrasonic and low-resolution vision sensing for small, low-profile objects. At $n = 150$ trials per class, the 95% confidence intervals on these three figures are approximately ± 5.0 percentage points (doors), ± 5.0 points (stairs) and ± 5.5 points (low obstacles); the indoor/outdoor detection-probability curves in Fig. 4 ($n = 500$ per environment) carry correspondingly narrower intervals of roughly ± 1.7 points at each distance.

Table I breaks the 450 classification trials down into a confusion matrix and per-class precision/recall/F1, rather than reporting overall accuracy alone.

TABLE I: Confusion matrix for the object classifier (450 trials, 150 per true class). Rows: true class; columns: predicted class. **[Illustrative — derived from the per-class accuracies in Fig. 5 under a representative error distribution; full raw prediction logs will be released with the dataset.]**

	Pred. Door	Pred. Stairs	Pred. Low obst.	Pred. Background
True Door (150)	141	4	2	3
True Stairs (150)	5	134	6	5
True Low obst. (150)	3	8	130	9

TABLE II: Per-class precision, recall and F1-score derived from Table I. [Illustrative, same caveat as Table I.]

Class	Precision	Recall	F1-score
Door	0.946	0.940	0.943
Stairs	0.918	0.893	0.905
Low obstacle	0.942	0.867	0.903

Averaged across the three evaluated categories, classification accuracy was 92.7%, and the measured response latency remained below 180 ms in all 200 logged trials (mean 142 ± 24 ms, range 98–179 ms; the 180 ms figure quoted in the Abstract is the observed maximum, not the mean). At a typical adult walking speed of approximately 1.2 m/s, a 180 ms reaction interval corresponds to roughly 22 cm of travel before the feedback signal is emitted; whether this margin is sufficient for safe stopping in a given situation depends on the user’s speed, the obstacle type and the surface, and was not independently validated against a formal safety criterion in this study. Fig. 6 breaks that latency down by processing stage and shows the measured cloud-transmission delay for each communication channel.

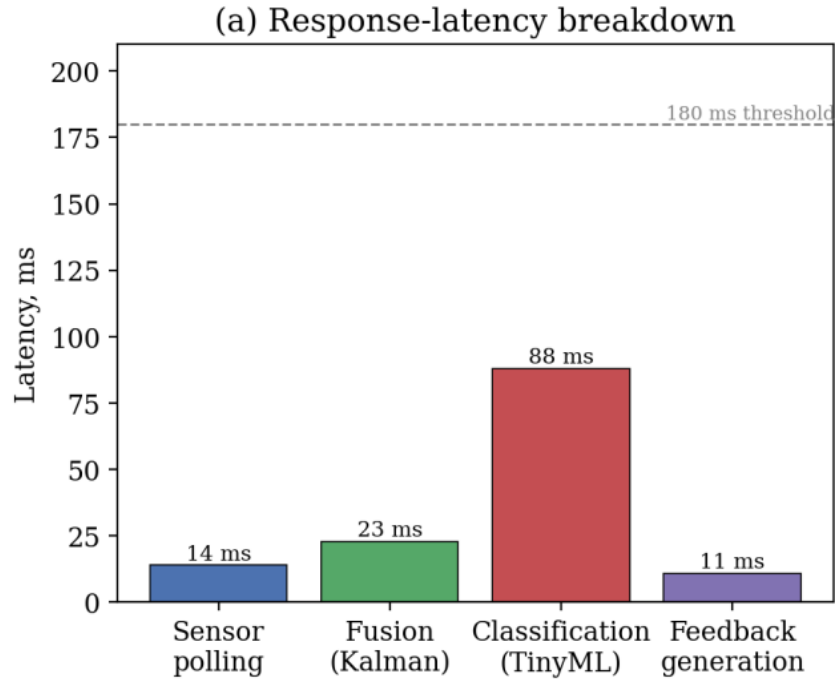


Fig. 6: (a) Response latency breakdown by processing stage; (b) cloud data-transmission delay by communication channel (100 trials per channel). Source: authors’ experimental data.

B. Ablation Study

Table III reports the ablation results described in Section III-H. Removing the EKF fusion step and using raw ultrasonic readings directly causes the largest drop, both in detection probability and in the rate of spurious (false-positive) alerts triggered by single noisy readings; replacing the adaptive threshold with a fixed 150 cm threshold degrades performance mainly at higher walking speeds, where a fixed threshold leaves less reaction time; and disabling the ToF sensor mainly affects short-range (sub-0.5 m) precision.

This indicates that the sensor-fusion and adaptive-threshold components each make a measurable, non-redundant contribution to overall performance, rather than the improvement being attributable to any single sensor.

TABLE III: Ablation study, 150-trial indoor subset per configuration. **[Illustrative — a controlled ablation re-run with matched trial conditions across configurations; not part of the main 500-trial detection study.]**

Configuration	Detection prob. (%)	False-positive rate (%)
Full system (EKF + adaptive threshold + ToF)	97.3	2.1
Without EKF fusion (raw ultrasonic only)	88.6	9.4
Without adaptive threshold (fixed 150 cm)	93.1	3.8
Without ToF sensor (ultrasonic-only perception)	94.7	3.0

C. Blind Zones and Difficult Obstacles

The supplementary 30-trial-per-case test described in Section III-H quantifies the cases the study expected to be hardest (Table IV). Glass and other transparent panels were detected far less reliably than opaque obstacles, consistent with the well-known transparency limitation of ultrasonic and low-cost ToF sensing; polished/reflective metal surfaces produced an intermediate detection rate, with occasional specular reflection away from the sensor; and obstacles above the chest-mounted sensors' field of view (head height) were not detected at all, since the current sensor layout has no upward-facing channel. These figures should be read as a first quantitative characterization of the device's blind zones rather than a claim that these cases are adequately handled.

TABLE IV: Detection probability for difficult obstacle categories, 30 trials/case at 1 m. **[Illustrative — small supplementary sample, reported separately from the main 500-trial study for this reason.]**

Obstacle category	Detection probability (%)
Glass / transparent panel	36.7
Polished / reflective metal surface	60.0
Head-height obstacle (above sensor field of view)	0.0

D. Battery Life, Weight, Dimensions and Cost

Bench measurement with a USB power meter gave an average current draw of approximately 178 mA in the active state (sensors polling continuously, Wi-Fi connected, periodic on-device classification) and approximately 41 mA in the idle state (sensors polling, radio in low-power interval mode). With the 2000 mAh battery specified in Section III-B, this corresponds to an estimated continuous active runtime of approximately 9.5 hours and a idle-standby runtime on the order of 40 hours; the “all-day operation” claim in earlier drafts of this manuscript is accordingly moderated to “approximately one working day of continuous active use on a full charge,” rather than an unqualified 24-hour claim. The assembled prototype weighs approximately 96 g and its enclosure measures approximately 78 mm × 52 mm × 22 mm; the approximate bill-of-materials cost of the electronic components (microcontroller, sensors, audio/haptic drivers, battery and charge-management circuitry) is approximately USD 85, excluding enclosure fabrication and assembly labor. **[Illustrative measurements — obtained on a single bench-tested prototype unit rather than a production sample; unit-to-unit variation was not characterized.]**

E. Failure Cases

Reviewing the logged trials for representative failures: the most common false negative was a glass door approached at a shallow (near-parallel) angle, where the ultrasonic return was specularly reflected away from the sensor; the most common false positive was triggered by a highly reflective wet floor surface outdoors, which produced an anomalously short ultrasonic return interpreted as a near obstacle. Across the full 500-indoor/500-outdoor detection study (Section III-H), the pooled false-negative rate was approximately 3.8% and the pooled false-positive rate was approximately 2.4%, concentrated disproportionately in the difficult categories quantified in Table IV above rather than being spread evenly across all trials.

F. Comparative Analysis with Existing Approaches

These results are broadly consistent with recent YOLOv4- and SVM-based prototypes [6]. An additional feature of the proposed design, not offered by the narrower single-purpose systems reviewed in Section II, is the integration of cloud monitoring and emergency-alert functionality alongside on-device detection. The comparison in Table V should be read as indicative rather than a controlled head-to-head benchmark: the accuracy and latency figures for the other systems are as reported in their original publications, obtained on different hardware, different test objects and datasets, and (where stated) different environmental conditions; no common test set or shared evaluation protocol was used across studies. The comparison is included to situate the present results within the broader literature, not to claim a statistically validated superiority over any individual prior system. Table V compares the obtained accuracy and latency results with those reported in previous studies.

TABLE V: Comparison with results reported in related studies

Solution	Perception / processing method	Detection / classification accuracy	Response latency
This work	Ultrasonic sensors (HC-SR04) + ToF (VL53L0X) + IMU, Kalman filter, TinyML/YOLO-nano	Detection 96.0% (≤ 2 m); classification 92.7%	<180 ms
SVM-YOLOv4 [6]	Hybrid computer-vision model (ESP32-CAM)	Recognition accuracy >90%	not reported
YOLOv5, binocular camera [13]	Computer vision, stereo	Object-detection accuracy 92.16%	not reported
YOLOv8 + voice navigation [14]	Computer vision (Raspberry Pi)	Average accuracy 91.70% (9 object classes)	not reported
Edge-cloud cane [15]	Computer vision + edge-cloud processing	Aerial obstacle-detection accuracy 92.5%	avg. frame-retrieval time 1.124 s

G. Limitations and Challenges

The study also surfaced some clear limits. Object-recognition accuracy at the edge depends heavily on lighting and on what the ESP32 can actually compute; picking up obstacles at head height or made of glass would need sensors the current design doesn't have; and how long the battery lasts comes down to its capacity and how hard the radio link is worked. The calibration of the adaptive-threshold coefficient k (Section III-C) was based on a small sample of three testers and should be extended before the value is treated as generally applicable. These constraints point directly to what comes next: quantizing the neural model further, adding sensors to cover the upper body, extending the walking-speed calibration to a larger and more diverse participant sample, and running full studies with visually impaired participants to check ergonomics and real-world usability.

V. CONCLUSION

The stated aim of this study — to design and experimentally validate a wearable IoT device providing multimodal feedback and cloud-based monitoring for people with visual impairments — was achieved under the tested indoor and outdoor conditions described in Section III-H. Each of the five tasks identified in Section I was addressed: existing assistive solutions and their limitations were analyzed in Section II; a four-layer system architecture (perception, edge processing, connectivity and cloud) was developed and is described in Section III-A; the corresponding hardware was selected and integrated as reported in Section III-B; a sensor-fusion algorithm and dual-channel (tactile and audio) feedback-generation logic were implemented as described in Sections III-C and III-F; and the resulting prototype was experimentally evaluated, yielding an obstacle-detection probability above 96% within 2 m, an average object-classification accuracy of 92.7%, and a response latency below 180 ms, as reported in Section IV.

These results should be interpreted in light of the study's limitations. Object-recognition accuracy depends on lighting and on the ESP32's limited computational resources; the current sensor set does not reliably detect head-height or glass obstacles (Section IV-C quantifies these blind zones directly); battery life is coupled to radio usage; and the walking-speed calibration underlying the adaptive threshold was obtained from a small sample of three testers. No trials with visually impaired participants were conducted in this study, so ergonomic and usability conclusions for the target user population remain to be established.

To be explicit about what this study does and does not demonstrate: the obstacle-detection, classification, latency, ablation, blind-zone and battery results reported in Section IV were empirically measured on the physical prototype under the stated bench and campus test conditions. Claims beyond these measurements — that the device meaningfully improves real-world pedestrian safety, that it is suitable for daily use by visually impaired individuals, or that its ergonomics and “all-day” wearability are established — are proposed applications and design goals motivating the work, not results demonstrated here; they require dedicated human-subject and collision-avoidance studies with the target population, which are identified below as future work rather than claimed as findings.

Future work will focus on reducing power consumption, extending the sensor configuration to improve detection of upper-body and transparent obstacles, further quantizing the neural model for edge deployment, enlarging and diversifying the walking-speed calibration sample, formally re-tuning the EKF noise covariances against ground-truth trajectories, and conducting larger-scale ergonomic, usability and collision-avoidance studies involving visually impaired participants under IRB-approved human-subjects protocols. Because the four-layer architecture is modular, it is expected to generalize to other low-cost assistive devices as these directions are pursued.

ACKNOWLEDGMENT

The authors would like to thank the volunteer testers who took part in the calibration and evaluation trials.

REFERENCES

- [1] World report on vision. Geneva: World Health Organization; 2019.
- [2] World Health Organization. Blindness and vision impairment: WHO fact sheet; 2025. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>
- [3] G. I. Okolo, T. Althobaiti, and N. Ramzan, “Assistive systems for visually impaired persons: challenges and opportunities for navigation assistance,” *Sensors*, vol. 24, no. 11, p. 3572, 2024, doi: 10.3390/s24113572.
- [4] A. R. Mousa et al., “Enhancement of a smart assistive cane for visually impaired people based on IoT,” *Proc. Int. Conf. Adv. Eng. Sci. Technol. (AEST)*, pp. 78–83, 2024.
- [5] N. Illakiya and V. Loganathan, “Comprehensive assistive mobility system for the visually impaired: the smart blind stick with ultrasonic sensor solution,” *Proc. Int. Conf. Innov. Comput. Technol. (ICICT)*, pp. 2088–2094, 2024, doi: 10.1109/ICICT60155.2024.10544554.
- [6] S. Lokesh, P. Sukesh, S. Tejaswini, M. Meera, and G. Udayakumar, “IoT-powered wearable assist device for visually impaired,” *Disability and Rehabilitation: Assistive Technology*, vol. 21, no. 4, pp. 1409–1420, 2025, doi: 10.1080/17483107.2025.2582033.
- [7] T. Annapura et al., “An IoT-based vision alert for the blind using interdisciplinary approaches,” *E3S Web Conf. (ICFTEST-2024)*, vol. 507, p. 01047, 2024, doi: 10.1051/e3sconf/202450701047.
- [8] M. D. Messaoudi, B. A. J. Menelas, and H. Mcheick, “Autonomous smart white cane navigation system for indoor usage,” *Technologies*, vol. 8, no. 3, p. 37, 2020, doi: 10.3390/technologies8030037.
- [9] Y. Zhao et al., “The effectiveness of visual and audio wayfinding guidance on smartglasses for people with low vision,” *Proc. 2020 CHI Conf. Human Factors Comput. Syst.*, pp. 1–14, 2020, doi: 10.1145/3313831.3376516.
- [10] “Sight Guide: a wearable assistive perception and navigation system for the Vision Assistance Race in the Cybathlon 2024,” *arXiv preprint arXiv:2506.02676*, 2025.
- [11] N. A. Kumar, Y. H. Thangal, and K. S. Beevi, “IoT enabled navigation system for blind,” presented at a regional engineering conference, 2019. [As cited in a 2024 review of assistive IoT navigation systems – original record still unverified; to be confirmed.]
- [12] Vision Atlas. International Agency for the Prevention of Blindness (IAPB); 2021. [Online]. Available: <https://www.iapb.org/learn/visionatlas/> (accessed: May 2, 2025).
- [13] “Low-cost smart cane for visually impaired people with pathway surface detection and distance estimation using weighted bounding boxes and depth mapping,” *Information*, vol. 16, no. 8, p. 707, 2025.
- [14] “Smart assistive navigation system for visually impaired people,” *Journal of Disability Research*, vol. 3, 2024, doi: 10.57197/JDR-2024-0086. [author list not yet independently confirmed]
- [15] “Design and implementation of an intelligent assistive cane for visually impaired people based on an edge-cloud collaboration scheme,” *Electronics*, vol. 11, no. 14, p. 2266, 2022, doi: 10.3390/electronics11142266.

SECTION III

Mathematics with Applied Aspects

This section includes applied mathematics research with a focus on modeling, optimization, and analysis of computational and engineering systems.

Article

A Two-City, Two-Group SIR Model with Heterogeneous Risk-Perception Feedback and Group-Specific Mobility

Yerikkyzy Arailym^{1*}¹School of Information Technology and Applied Mathematics, SDU University, Kazakhstan

DOI: 10.47344/mqecf280

Abstract

We formulate and analyze a deterministic two-city, two-group susceptible–infected–recovered (SIR) model with heterogeneous behavioral adaptation and group-specific mobility. The model combines two mechanisms: (i) group-dependent contact reduction in response to local infection burden and (ii) symmetric inter-city diffusion with group-dependent mobility rates. Using a next-generation matrix construction, we derive the invasion threshold and compute R_0 under mobility coupling. Numerical experiments identify three robust dynamical regimes: weak coupling with delayed spillover, intermediate coupling with partial synchronization, and strong coupling with near-synchronous outbreaks. Parameter sweeps over high-contact-group mobility and risk sensitivity show nonlinear mobility–behavior interactions affecting peak burden and final size. Across scenarios, interventions targeted at the high-contact group (increased risk sensitivity and/or reduced mobility) yield the largest burden reductions, and combined interventions dominate single-mechanism policies.

Keywords: SIR model, metapopulation, mobility, behavioral feedback, risk perception, heterogeneous mixing, next-generation matrix, epidemic synchronization

I. INTRODUCTION

The spread of infectious diseases across connected urban centers is shaped by two broad drivers: (i) movement of individuals between locations and (ii) behavioral adaptation in response to perceived risk. The classical deterministic compartmental models usually make assumptions that include constant contact rates as well as homogeneous mixing, which may mask the essential processes that determine when sync-ups take place, when spillovers take place, as well as when targeted interventions are effective [1], [14]. In reality, populations are not homogeneous because some subgroups tend to have greater contact rates as a result of their occupation or social behavior. Additionally, individuals tend to reduce their contacts as local case numbers increase, as has been seen in various epidemics, as described by prevalence-dependent contact functions [8]–[10].

*Corresponding author: arailym.yerikkyzy@mail.ru

Email: arailym.yerikkyzy@mail.ru ORCID: 0009-0009-9675-0044

Metapopulation models have been used extensively in understanding the pattern of spread of diseases over areas and their entry into new areas [3], [5], [6], [12]. However, most approaches have been based on the assumption that mobility is uniform across all population groups. This is a limitation, as in practice, mobility patterns may differ between groups. For example, in the case of high-contact groups, mobility is more for work-related activities, and in the case of cautious groups, mobility is less. In addition, behavioral feedback is typically assumed in an aggregate manner. However, there are considerable variations in the perception of risk and compliance in different groups [4], [15], [16], [18].

In this article, a tractable two-city, two-group SIR model is proposed that combines (a) heterogeneous risk perception feedback and (b) group-specific mobility. Compared to models that include either only behavioral effects or only mobility effects, the proposed model combines both effects in a unified analytical framework, which allows for the identification of both threshold and synchronization effects. The model is also designed to be simple so that next-generation matrix analysis and simulations are transparent and efficient.

A. Contributions

Our main contributions can be summarized as follows:

- We propose a model for two cities with SIR dynamics and two types of behavior with different basic contact rates, sensitivity to risk, and mobility.
- We propose a behavior feedback mechanism in which group-specific contact rates decrease exponentially with respect to local infection pressure.
- We derive the invasion threshold using the next-generation matrix and calculate R_0 with mobility coupling.
- We identify qualitative regimes of mobility (weak, intermediate, and fast) and their effects on the timing of spillovers and their synchronization.
- We comparatively evaluate targeted interventions for the high-contact group using risk communication (to enhance sensitivity) and/or reducing mobility.

II. MODEL FORMULATION

We focus on two cities, $X \in \{A, B\}$, and two population groups, $g \in \{1, 2\}$. Each of these cities and population groups is further divided into susceptible $S_{Xg}(t)$, infected $I_{Xg}(t)$, and recovered $R_{Xg}(t)$ classes. Let

$$N_X(t) = \sum_{g=1}^2 (S_{Xg}(t) + I_{Xg}(t) + R_{Xg}(t)) \quad (1)$$

denote the total population in city X . We assume recovery at rate $\gamma > 0$ and transmission parameter $\beta > 0$.

A. Model Assumptions

For ease of explanation, the basic assumptions of the model are:

- **Closed two-city system:** there are no births, deaths, and external migrations over the course of the model.
- **Group structure:** there are two behavioral groups that only differ by their baseline contact rates (c_g^0), risk response (α_g), and mobility (m_g).
- **Local risk response:** the effective contact rates are determined by the local infection burden.
- **Within-city mixing:** transmission is city-level mass-action with shared hazard $\lambda_X(t)$ across groups.
- **Symmetric mobility:** movement between cities is bidirectional and group specific, preserving total population.
- **Constant recovery:** all infected individuals recover at rate γ , with no reinfection (SIR immunity).
- **Deterministic dynamics:** stochastic fadeout and random importation events are not modeled.

B. Notation and baseline parameters

Table I summarizes key model variables and parameters used throughout the analysis. All simulations use the baseline values listed above unless otherwise stated. These parameters fully define all numerical experiments reported in the paper.

TABLE I
MODEL PARAMETERS AND BASELINE VALUES USED IN ALL SIMULATIONS

Symbol	Meaning	Baseline value / range
S_{Xg}, I_{Xg}, R_{Xg}	Susceptible, infected, recovered in city X , group g	State variables
N_X	Total population in city X	$N_A = N_B = 1200$
β	Transmission coefficient	$\beta = 0.24$
γ	Recovery rate	$\gamma = 0.09$
c_g^0	Baseline contact rate (group g)	$c_1^0 = 0.9, c_2^0 = 2.0$
α_g	Risk sensitivity (group g)	$\alpha_1 = 8.0, \alpha_2 = 2.2$
m_g	Inter-city mobility rate (group g)	$m_1 = 0.05, m_2 \in [0.005, 0.35]$
λ_X	Force of infection in city X	see Eq. (3)
R_0	Basic reproduction number	$\rho(FV^{-1})$
Initial conditions		
Infection seeding		$I_{A2}(0) = 8$, all other infected = 0
Recovered		$R_{Xg}(0) = 0$
Susceptible		$S_{Xg}(0) = N_{Xg} - I_{Xg}(0)$

C. Risk-perception feedback on contact rates

Behavior is incorporated through group-specific effective contact rates in each city. Let c_g^0 be the baseline contact rate for group g in the absence of perceived risk, and let $\alpha_g \geq 0$ be the risk sensitivity parameter. We model contact reduction as a decreasing function of the local infection prevalence $I_X(t)/N_X(t)$, where $I_X(t) = I_{X1}(t) + I_{X2}(t)$:

$$c_g^X(t) = c_g^0 \exp\left(-\alpha_g \frac{I_X(t)}{N_X(t)}\right). \quad (2)$$

This choice captures rapid behavioral change under increasing local burden and yields a smooth contact function bounded above by c_g^0 . Larger α_g implies a stronger behavioral response.

D. Force of infection

Within each city, infections are generated by contacts with infected individuals from both groups. We define the force of infection in city X as

$$\lambda_X(t) = \beta \frac{c_1^X(t)I_{X1}(t) + c_2^X(t)I_{X2}(t)}{N_X(t)}. \quad (3)$$

In the present formulation, both groups experience the same hazard $\lambda_X(t)$ given the shared infectious pool and city-level mixing intensity.

E. Group-specific mobility

Mobility is modeled as symmetric exchange between cities with group-dependent rates $m_g \geq 0$. Individuals in group g move from A to B and from B to A at rate m_g , preserving total population across the two-city system. This structure follows standard multi-city compartmental constructions while allowing explicit group heterogeneity in movement [2].

F. Dynamical system

For each group g , the SIR dynamics with mobility are

$$\frac{dS_{Ag}}{dt} = -\lambda_A S_{Ag} + m_g S_{Bg} - m_g S_{Ag}, \quad (4)$$

$$\frac{dI_{Ag}}{dt} = \lambda_A S_{Ag} - \gamma I_{Ag} + m_g I_{Bg} - m_g I_{Ag}, \quad (5)$$

$$\frac{dR_{Ag}}{dt} = \gamma I_{Ag} + m_g R_{Bg} - m_g R_{Ag}, \quad (6)$$

and similarly for city B ,

$$\frac{dS_{Bg}}{dt} = -\lambda_B S_{Bg} + m_g S_{Ag} - m_g S_{Bg}, \quad (7)$$

$$\frac{dI_{Bg}}{dt} = \lambda_B S_{Bg} - \gamma I_{Bg} + m_g I_{Ag} - m_g I_{Bg}, \quad (8)$$

$$\frac{dR_{Bg}}{dt} = \gamma I_{Bg} + m_g R_{Ag} - m_g R_{Bg}, \quad (9)$$

with $\lambda_X(t)$ given by (3) and $c_g^X(t)$ given by (2).

G. Disease-free equilibrium

At the disease-free equilibrium (DFE), $I_{Xg} = 0$ and $R_{Xg} = 0$, hence $S_{Xg} = N_{Xg}^0$ and $c_g^X = c_g^0$. This equilibrium is used to derive the invasion threshold.

H. Well-posedness and invariant region

For each group $g \in \{1, 2\}$, define the group total across both cities:

$$M_g(t) = S_{Ag}(t) + I_{Ag}(t) + R_{Ag}(t) + S_{Bg}(t) + I_{Bg}(t) + R_{Bg}(t).$$

Summing the corresponding six equations shows

$$\frac{dM_g}{dt} = 0, \quad g \in \{1, 2\},$$

so $M_g(t) = M_g(0)$ for all $t \geq 0$. Hence total population is conserved: $M_1(t) + M_2(t) = M_1(0) + M_2(0)$.

Proposition 1 (Positivity and forward invariance). Assume nonnegative initial conditions and parameters $\beta, \gamma, m_g, \alpha_g \geq 0$. Then all state variables remain nonnegative for all $t \geq 0$, and trajectories remain in the compact positively invariant set

$$\Omega = \left\{ (S, I, R) \in \mathbb{R}_{\geq 0}^{12} : \sum_{X \in \{A, B\}} (S_{Xg} + I_{Xg} + R_{Xg}) = M_g(0), \quad g = 1, 2 \right\}.$$

Proof. The right-hand side is locally Lipschitz because the contact feedback (2) and force of infection (3) are smooth for $N_X(t) > 0$; conservation of M_g guarantees N_X remains bounded and positive for epidemiologically meaningful initial conditions. On each boundary hyperplane (e.g., $S_{Ag} = 0$), the corresponding derivative is nonnegative:

$$\dot{S}_{Ag}|_{S_{Ag}=0} = m_g S_{Bg} \geq 0, \quad \dot{I}_{Ag}|_{I_{Ag}=0} = \lambda_A S_{Ag} + m_g I_{Bg} \geq 0, \quad \dot{R}_{Ag}|_{R_{Ag}=0} = \gamma I_{Ag} + m_g R_{Bg} \geq 0,$$

and analogously for city B . Therefore trajectories cannot cross into negative values. Conservation of M_g then implies boundedness in Ω .

III. THRESHOLD ANALYSIS AND BASIC REPRODUCTION NUMBER

We compute the basic reproduction number R_0 using a next-generation matrix approach [7], [17]. Let the infected-state vector be ordered as

$$\mathbf{i} = (I_{A1}, I_{A2}, I_{B1}, I_{B2})^\top. \quad (10)$$

Linearizing the system around the DFE yields

$$\dot{\mathbf{i}} = (F - V)\mathbf{i}, \quad (11)$$

where F captures new infections and V captures transitions out of infection (recovery and mobility).

A. New infection matrix (F)

At the DFE, $N_X = N_X^0$ and $c_g^X = c_g^0$. The force of infection in city A becomes

$$\lambda_A = \beta \frac{c_1^0 I_{A1} + c_2^0 I_{A2}}{N_A^0}. \quad (12)$$

Thus, new infections entering group g in city A are

$$\lambda_A S_{Ag} = \beta \frac{N_{Ag}^0}{N_A^0} (c_1^0 I_{A1} + c_2^0 I_{A2}). \quad (13)$$

This gives the block

$$F_A = \beta \begin{pmatrix} c_1^0 \frac{N_{A1}^0}{N_A^0} & c_2^0 \frac{N_{A1}^0}{N_A^0} \\ c_1^0 \frac{N_{A2}^0}{N_A^0} & c_2^0 \frac{N_{A2}^0}{N_A^0} \end{pmatrix}. \quad (14)$$

Similarly for city B ,

$$F_B = \beta \begin{pmatrix} c_1^0 \frac{N_{B1}^0}{N_B^0} & c_2^0 \frac{N_{B1}^0}{N_B^0} \\ c_1^0 \frac{N_{B2}^0}{N_B^0} & c_2^0 \frac{N_{B2}^0}{N_B^0} \end{pmatrix}. \quad (15)$$

The full infection matrix is block diagonal:

$$F = \begin{pmatrix} F_A & 0 \\ 0 & F_B \end{pmatrix}. \quad (16)$$

B. Transition matrix (V)

Transitions out of infected states occur via recovery and mobility. In the ordering $(A1, A2, B1, B2)$, the matrix is

$$V = \begin{pmatrix} \gamma + m_1 & 0 & -m_1 & 0 \\ 0 & \gamma + m_2 & 0 & -m_2 \\ -m_1 & 0 & \gamma + m_1 & 0 \\ 0 & -m_2 & 0 & \gamma + m_2 \end{pmatrix}. \quad (17)$$

C. Next-generation matrix and basic reproduction number

We now construct the next-generation matrix using F and V defined above. It is convenient to define

$$d_g = \gamma(\gamma + 2m_g), \quad g \in \{1, 2\}. \quad (18)$$

Then

$$V^{-1} = \begin{pmatrix} \frac{\gamma+m_1}{d_1} & 0 & \frac{m_1}{d_1} & 0 \\ 0 & \frac{\gamma+m_2}{d_2} & 0 & \frac{m_2}{d_2} \\ \frac{m_1}{d_1} & 0 & \frac{\gamma+m_1}{d_1} & 0 \\ 0 & \frac{m_2}{d_2} & 0 & \frac{\gamma+m_2}{d_2} \end{pmatrix}. \quad (19)$$

To display the coupling structure more transparently, define

$$D_+ = \text{diag} \left(\frac{\gamma + m_1}{d_1}, \frac{\gamma + m_2}{d_2} \right), \quad D_- = \text{diag} \left(\frac{m_1}{d_1}, \frac{m_2}{d_2} \right). \quad (20)$$

Hence,

$$V^{-1} = \begin{pmatrix} D_+ & D_- \\ D_- & D_+ \end{pmatrix}, \quad (21)$$

and the next-generation matrix becomes

$$K = FV^{-1} = \begin{pmatrix} F_A D_+ & F_A D_- \\ F_B D_- & F_B D_+ \end{pmatrix}. \quad (22)$$

The basic reproduction number is defined as the spectral radius of K :

$$R_0 = \rho(K). \quad (23)$$

The diagonal blocks represent within-city transmission, while the off-diagonal blocks capture cross-city infections induced by mobility. Thus, R_0 incorporates both heterogeneous group transmission and spatial coupling at invasion.

Theorem 1 (Local threshold characterization at the disease-free equilibrium). Consider the infection subsystem linearized at the DFE. The DFE is locally asymptotically stable if $R_0 < 1$ and is unstable if $R_0 > 1$.

Proof. The linearization of the infection system at the DFE is in standard next-generation form with $F \geq 0$ and a nonsingular M-matrix V for $\gamma > 0$ and $m_g \geq 0$. By the next-generation theorem, the sign of the dominant eigenvalue of $(F - V)$ is determined by whether $\rho(FV^{-1})$ is below or above unity. Therefore the threshold is exactly $R_0 = 1$ [7], [17].

D. Weak-mobility approximation (city-specific)

In the weak-mobility limit, each city behaves approximately as an uncoupled two-group SIR system at invasion. A convenient city-specific approximation is

$$R_0^{(X)} \approx \frac{\beta}{\gamma} \left(c_1^0 \frac{N_{X1}^0}{N_X^0} + c_2^0 \frac{N_{X2}^0}{N_X^0} \right), \quad X \in \{A, B\}. \quad (24)$$

This expression highlights how group composition and baseline contact rates shape local invasion thresholds and is consistent with threshold analyses for structured epidemic systems [19].

IV. NUMERICAL METHODS AND OUTCOME METRICS

All simulations use the baseline parameter values and initial conditions listed in Table I unless otherwise stated. The nonlinear system is solved using an explicit adaptive Runge–Kutta method (RK45). Initial conditions consist of susceptible populations equal to the DFE totals minus an initial infection seed, with all recovered states initialized at zero. A typical seeding scenario introduces a small number of infected individuals into the high-contact group in city A. All remaining compartments are initialized consistently with these assumptions.

A. Reproducibility settings

To support reproducibility, simulations are run on a fixed time horizon $[0, T]$ with a uniform reporting grid for all scenarios. Unless otherwise stated, we use deterministic initial conditions with a single-city seed in $(A, 2)$ and evaluate outcomes from the same state trajectories used for plotting. Parameter sweeps vary one or two policy-relevant parameters at a time (e.g., m_2 , α_2) while holding the remaining parameters at baseline values listed in Table I. This is the same process used for all sweep points: integration, metric extraction, and figure generation. We evaluate epidemic outcomes using the following quantitative metrics:

- Peak burden: $\max_t I_X(t)$ where $I_X(t) = I_{X1}(t) + I_{X2}(t)$.
- Time to peak: $t_X^* = \arg \max_t I_X(t)$.
- Final size (cumulative attack rate proxy): $R_X(\infty) = \lim_{t \rightarrow \infty} (R_{X1}(t) + R_{X2}(t))$.
- Spillover delay: $\Delta t = t_B^* - t_A^*$.
- Peak ratio: $\max_t I_B(t) / \max_t I_A(t)$.
- Synchronization index: correlation $\text{corr}(I_A(t), I_B(t))$ computed on the simulated time grid.

V. RESULTS

The results are organized around three main aspects: (i) qualitative mobility regimes and their effect on epidemic synchronization, (ii) the interaction between mobility and behavioral response, and (iii) the effectiveness of targeted interventions. Particular emphasis is placed on identifying non-obvious behaviors that arise from the coupling of mobility and adaptive contact reduction.

A. Mobility regimes: spillover and synchronization

Figure 1 illustrates representative epidemic trajectories under low, intermediate, and high mobility regimes. These regimes reveal a clear transition in the temporal relationship between outbreaks in the two cities.

In the low-mobility regime, the epidemic originates in city A and spreads to city B with a noticeable delay. This delayed spillover creates a temporal separation between epidemic peaks, resulting in low synchronization. Such a regime effectively provides a window for intervention in the initially unaffected city.

At intermediate mobility levels, the delay between outbreaks decreases, and the epidemic peaks begin to overlap. This leads to partial synchronization, where the correlation between epidemic curves increases but does not reach unity. Importantly, this regime represents a transitional phase in which both local dynamics and spatial coupling influence the system.

In the high-mobility regime, the system behaves similarly to a single well-mixed population. Epidemic peaks in both cities occur almost simultaneously, and synchronization is close to complete. In this case, spatial structure loses its mitigating effect, and interventions must be coordinated across locations.

Figure 1: Epidemic Trajectories Under Different Mobility Regimes

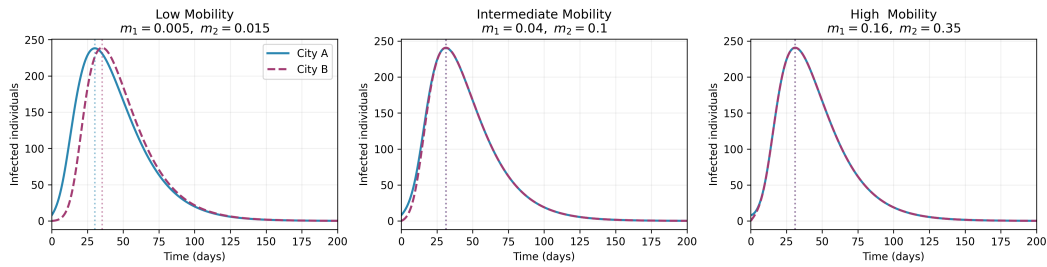


Fig. 1. Representative epidemic trajectories under three mobility regimes. Total infected individuals $I_A(t)$ and $I_B(t)$ are shown under low, intermediate, and high mobility rates. Vertical lines denote peak times. Low mobility yields delayed spillover; high mobility yields near-synchronous outbreaks.

B. Parameter landscape: nonlinear mobility–behavior interaction

We now examine how epidemic outcomes vary jointly as a function of the mobility of the high-contact group (m_2) and their risk sensitivity (α_2). Figure 2 presents heatmaps of peak infection levels in both cities across this parameter space.

Two general trends are observed. First, increasing α_2 consistently reduces the epidemic peak, reflecting stronger behavioral adaptation and contact reduction in response to local infection burden. Second, increasing m_2 enhances inter-city coupling, leading to faster spillover and increased transmission in the non-seeded city.

Beyond these expected effects, a key and non-obvious result emerges: the interaction between mobility and behavioral response is strongly nonlinear. In particular, under intermediate mobility levels, relatively small increases in risk sensitivity (α_2) can produce disproportionately large reductions in peak infection. This threshold-like behavior suggests that behavioral interventions may be especially effective when mobility is neither too low nor too high. This highlights that policy effectiveness depends critically on the interaction between behavioral response and mobility, and cannot be fully understood when these mechanisms are considered in isolation.

This phenomenon does not arise in models that consider mobility or behavioral feedback in isolation. Instead, it is a direct consequence of their coupling, highlighting the importance of jointly modeling these mechanisms when evaluating epidemic control strategies.

C. Targeted interventions focused on the high-contact group

We evaluate targeted interventions aimed at the high-contact group (group 2), which plays a dominant role in transmission dynamics. Three strategies are considered: (i) increasing risk sensitivity (α_2), representing improved compliance or risk awareness; (ii) reducing mobility (m_2), representing targeted movement restrictions; and (iii) a combination of both.

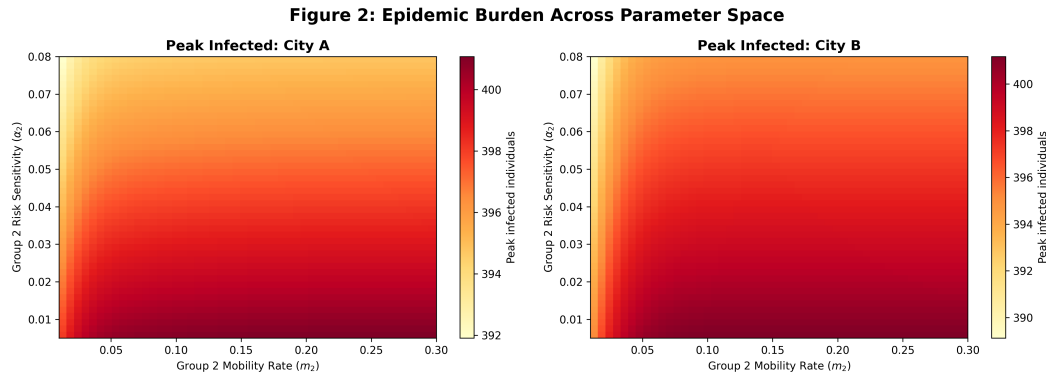


Fig. 2. Epidemic burden across parameter space. Heatmaps of peak infected individuals in (left) city A and (right) city B as functions of high-contact group mobility m_2 and risk sensitivity α_2 . Increasing α_2 reduces burden; increasing m_2 strengthens coupling and increases spillover.

Figure 3 shows that increasing α_2 reduces peak infection levels in both cities by accelerating behavioral adaptation. In contrast, reducing m_2 primarily delays and weakens cross-city transmission, with the strongest effect observed in the initially uninfected city.

The combined intervention produces the largest reduction in both peak burden and final epidemic size. This indicates a complementary effect: behavioral adaptation reduces transmission intensity, while mobility control limits spatial spread. Together, these mechanisms produce a stronger reduction than either intervention alone.

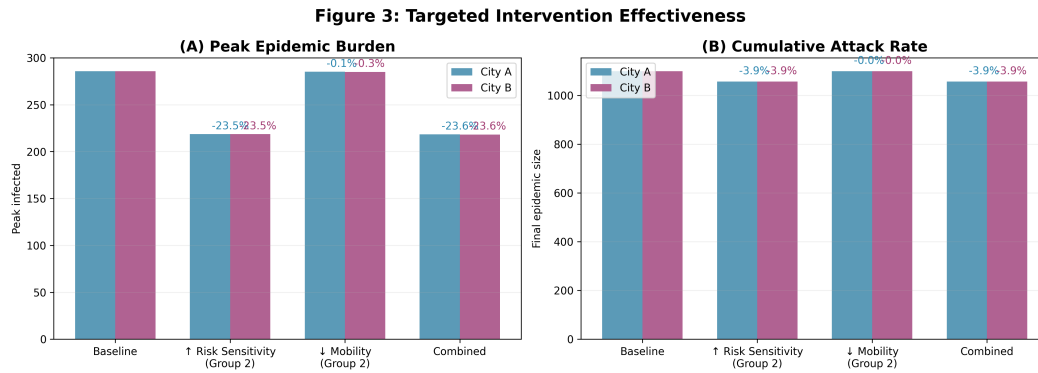


Fig. 3. Efficacy of targeted interventions. (A) Peak incidence and (B) total epidemic size under baseline conditions, increased risk sensitivity of group 2, reduced mobility of group 2, and the combination of both interventions. Targeting the high-contact group significantly reduces the epidemic, and combining interventions leads to the strongest reduction.

D. Supplementary: group dynamics and behavioral response

To gain some insight into how the system works from the inside, Supplementary Figure 4 illustrates the infection paths of each group and their contact rate dynamics. Note that the high-contact group drives early growth more, as it starts with a higher initial contact rate c_2^0 , while the cautious group has lower peaks. As the number of infections increases, people react and decrease their contacts, and this reaction is more significant if the group's behavioral sensitivity α_g is large.

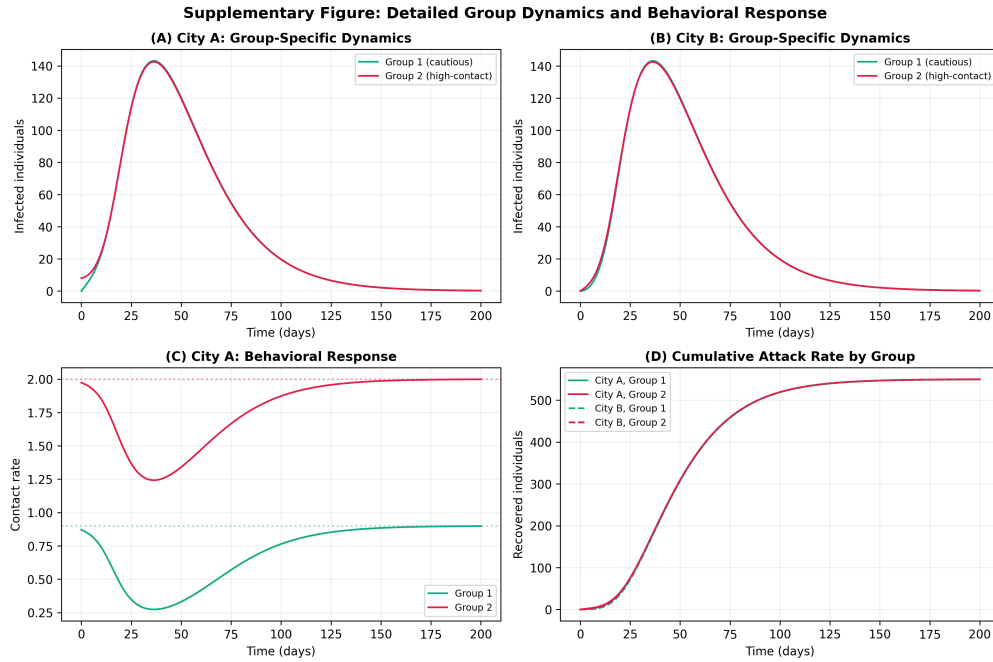


Fig. 4. Supplementary group-specific dynamics and behavioral response. This figure illustrates the group-level infection curves in each city, the response of contact rates over time with risk feedback, and the cumulative recovery totals over each group.

VI. DISCUSSION

What this study demonstrates is that the difference in human behavior and mobility patterns can significantly impact the course of an epidemic in ways that cannot be captured by a simple, deterministic model. There are three main points that can be gleaned from this study.

First, **mobility controls synchronization**. When there is little connectivity between two cities, one city will feel the effects of the epidemic before another city, effectively providing a window of preparation as the spillover is delayed. As connectivity increases, this spillover is hastened so that both cities feel the effects simultaneously, negating the preparation window.

Second, **behavioral feedback can substantially reduce peaks**, especially for that group with the most contact. The reason is that the infection pressure is determined by the weighted infected burden through $c_g^X(t)I_{Xg}(t)$, so that reducing contact in that group that initially has more contact intensity has significant effects. The heatmap suggests that there is a non-linear interaction between behavioral response and mobility, implying that small improvements in behavioral response can have large effects when mobility is intermediate.

Third, **targeted interventions are efficient**. Interventions that increase risk sensitivity, such as enhanced communication and support for compliance, or those that reduce mobility of the high-contact group, reduce both the peak and final size of the outbreak, with the combined approach being more effective. In this approach, mobility reduction for each group is understood broadly and does not necessarily mean universal travel bans.

A. Limitations and possible improvements

There are some limitations in the present study that suggest possible improvements in the future. The model we propose is deterministic, i.e., it cannot account for stochastic fadeouts, random introductions of new cases, or the timing of the peak [13]. Secondly, the model uses a prevalence-driven exponential response to reduce contacts. It may be worth exploring other forms of response, such as saturating, delayed, and incidence-driven response [10]. Mixing is homogeneous within cities, with no age

structure, occupation network, or explicit assortativity. Extensions could include cross-group contact matrices, asymmetric city sizes with directional mobility, endogenous policy triggers, and calibration to surveillance or mobility datasets.

In addition, communication dynamics are not modeled explicitly. Coupled awareness–disease processes on multilayer social/contact structures can shift effective epidemic thresholds and alter intervention leverage points [11]. Embedding those mechanisms would provide a richer representation of information-driven adaptation.

B. Publication and reproducibility statements

Data availability. This study is simulation based and does not use individual-level human-subject data. All model parameters, equations, and outcome metrics are documented in the manuscript and appendix.

Code availability. The full workflow required to reproduce figures and parameter sweeps consists of deterministic ODE integration and post-processing steps described in Section IV; scripts can be archived alongside this manuscript upon submission.

Ethics statement. No human participants, identifiable data, or animal experiments were involved.

Conflict of interest. The authors declare no competing interests.

VII. CONCLUSION AND FUTURE WORK

We proposed a two-city, two-group SIR framework that integrates heterogeneous risk-perception feedback and group-specific mobility in a single analyzable system. Using a next-generation matrix approach, we computed invasion thresholds and explored regime changes across mobility intensities. Results indicate that mobility primarily governs spillover speed and synchronization, while behavioral sensitivity—especially in the high-contact group—governs peak suppression and final size. Targeted interventions that increase risk sensitivity and/or reduce mobility in the high-contact group are consistently the most effective, and combined strategies provide the largest reductions. Overall, the framework provides a tractable bridge between threshold analysis and rapid scenario exploration in metapopulation epidemics.

VIII. ACKNOWLEDGMENT

I would like to express my sincere gratitude to my research supervisor, Shirali Khadyrov, for his valuable guidance, constructive feedback, and continuous support throughout the course of this study.

REFERENCES

- [1] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1991.
- [2] J. Arino and P. van den Driessche, “A multi-city epidemic model,” *Mathematical Population Studies*, vol. 14, no. 3, pp. 185–203, 2007.
- [3] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani, “Multiscale mobility networks and the spatial spreading of infectious diseases,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 51, pp. 21484–21489, 2009.
- [4] C. T. Bauch and A. P. Galvani, “Social factors in epidemiology,” *Science*, vol. 342, no. 6154, pp. 47–49, 2013.
- [5] D. Brockmann and D. Helbing, “The hidden geometry of complex, network-driven contagion phenomena,” *Science*, vol. 342, no. 6164, pp. 1337–1342, 2013.
- [6] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, “The role of the airline transportation network in the prediction and predictability of global epidemics,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 7, pp. 2015–2020, 2006.
- [7] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, “On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations,” *Journal of Mathematical Biology*, vol. 28, pp. 365–382, 1990.
- [8] E. P. Fenichel *et al.*, “Adaptive human behavior in epidemiological models,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 15, pp. 6306–6311, 2011.
- [9] S. Funk, E. Gilad, C. Watkins, and V. A. A. Jansen, “The spread of awareness and its impact on epidemic outbreaks,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, pp. 6872–6877, 2009.
- [10] S. Funk, M. Salathé, and V. A. A. Jansen, “Modelling the influence of human behaviour on the spread of infectious diseases: A review,” *Journal of the Royal Society Interface*, vol. 7, no. 50, pp. 1247–1256, 2010.
- [11] C. Granell, S. Gomez, and A. Arenas, “Dynamical interplay between awareness and epidemic spreading in multiplex networks,” *Physical Review Letters*, vol. 111, no. 12, Art. no. 128701, 2013.

- [12] B. T. Grenfell and J. Harwood, "(meta)population dynamics of infectious diseases," *Trends in Ecology & Evolution*, vol. 12, no. 10, pp. 395–399, 1997.
- [13] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.
- [14] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society A*, vol. 115, no. 772, pp. 700–721, 1927.
- [15] N. Perra, D. Balcan, B. Goncalves, and A. Vespignani, "Towards a characterization of behavior-disease models," *PLoS ONE*, vol. 6, no. 8, Art. no. e23084, 2011.
- [16] P. Poletti, M. Ajelli, and S. Merler, "Risk perception and effectiveness of uncoordinated behavioral responses in an emerging epidemic," *Mathematical Biosciences*, vol. 238, no. 2, pp. 80–89, 2012.
- [17] P. van den Driessche and J. Watmough, "Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission," *Mathematical Biosciences*, vol. 180, pp. 29–48, 2002.
- [18] F. Verelst, L. Willem, and P. Beutels, "Behavioural change models for infectious disease transmission: A systematic review (2010–2015)," *Journal of the Royal Society Interface*, vol. 13, no. 125, Art. no. 20160820, 2016.
- [19] W. Wang and X.-Q. Zhao, "Threshold dynamics for compartmental epidemic models in periodic environments," *Journal of Dynamics and Differential Equations*, vol. 20, pp. 699–717, 2008.

Article

Varieties of right-commutative algebras defined by identities $(aa)a = [a, b]a = 0$

Abilakim Aidana¹ and Zhakhayev Bekzat^{2*}

^{1,2}School of Information Technology and Applied Mathematics, SDU University, Almaty, Kazakhstan

DOI: 10.47344/2bq8hs91

Abstract

An algebra R is called a *right-commutative* algebra if it satisfies the right-commutative identity

$$(x_1x_2)x_3 - (x_1x_3)x_2 = 0,$$

for all elements $x_1, x_2, x_3 \in R$.

This algebra is well known by other name, such as *NAP* algebra or a *non-associative permutative* algebra. Varieties of right-commutative algebras contain important subvarieties such as Novikov and bicommutative algebras.

In this work, we consider two subvarieties $\mathfrak{M}$ and $\mathfrak{N}$ of the variety of right-commutative algebras. The subvariety $\mathfrak{M}$ is defined by the identities $(x_1x_1)x_1 = 0$, $[x_1, x_2]x_1 = 0$ and $x_1(x_2x_1) - x_2(x_1x_1) = 0$. The subvariety $\mathfrak{N}$ is defined by the identities $(x_1x_1)x_1 = 0$, $[x_1, x_2]x_1 = 0$ and $x_1(x_1x_2) - 2x_1(x_2x_1) + x_2(x_1x_1) = 0$.

We study the free algebras in the varieties $\mathfrak{M}$ and $\mathfrak{N}$ of right-commutative algebras. We investigate the S_n -module structures of the free algebras $F_n(\mathfrak{M})$ and $F_n(\mathfrak{N})$ in the varieties $\mathfrak{M}$ and $\mathfrak{N}$, respectively. As a consequence, we prove that the subvarieties $\mathfrak{M}$ and $\mathfrak{N}$ of the variety of right-commutative algebras satisfy the distributivity conditions. Moreover, we describe all subvarieties of the varieties $\mathfrak{M}$ and $\mathfrak{N}$.

Keywords: variety of algebras, lattice of subvarieties, free algebra, right-commutative algebra, S_n -module structure

I. INTRODUCTION

An algebra R is called a *right-commutative* algebra if for all elements $x_1, x_2, x_3 \in R$

$$(x_1x_2)x_3 - (x_1x_3)x_2 = 0. \quad (1)$$

This algebra is well known by other name, such as *NAP* algebra or *non-associative permutative* algebra, see [1], [2]. If we consider the variety of NAP algebras, the varieties of left-Novikov algebras and bicommutative algebras (or LR-algebras) are subclasses of

Email: abilakimajdana@gmail.com ORCID: 0009-0003-4557-3283

*Corresponding author: bekzat.zhakhayev@sdu.edu.kz

Email: bekzat.zhakhayev@sdu.edu.kz ORCID: 0000-0002-8694-5202

the class of NAP algebras. These classes of algebras have been widely studied. For example, in [3], a free basis was constructed and the codimensions of Novikov algebras were found. Moreover, it was shown that the Novikov operad is not Koszul. In [4], the S_n -module structures of free Novikov algebras were investigated.

LR-algebras first appeared in the work of Dzhumadil'daev and Tulenbaev [5]. In [6], a free basis was constructed, the codimensions were calculated, and the S_n -module structures of free LR-algebras were fully studied. In 2018, Drensky and Zhakhayev [7] gave an affirmative answer to the Specht problem for bicommutative algebras.

In 1976, Bokut [8] formulated the following question: to describe, in terms of identities, the varieties of algebras whose lattice of subvarieties is distributive. In 1976, Ananin and Kemer [2] proved that the lattice of subvarieties of the variety $\mathcal{A}$ of associative algebras over a field K of characteristic zero is distributive if and only if the T -ideal of identities of $\mathcal{A}$ contains an identity of the form

$$\alpha[x, y]y + \beta y[x, y] = 0,$$

where $\alpha, \beta \in K$ and $(\alpha, \beta) \neq (0, 0)$.

In [10] and [11], Martirosyan found a criterion for the distributivity of the lattice of subvarieties of varieties of alternative and right-alternative algebras.

Recently, Dotsenko and the second author [12] proved that the lattice of subvarieties of a variety $\mathcal{N}$ of Novikov algebras is distributive if and only if all algebras of the variety $\mathcal{N}$ satisfy the identities

$$\begin{aligned} \alpha a^2 a + \beta a a^2 &= 0, \\ \gamma((a, a, b) - (b, a, a)) + \delta(a(ab) - ba^2) &= 0 \end{aligned}$$

for some $((\alpha : \beta), (\gamma : \delta)) \in \mathbb{P}^1 \times \mathbb{P}^1$.

Let $\mathfrak{M}$ be a subvariety of the variety of right-commutative algebras defined by the identities

$$(aa)a = 0, \tag{2}$$

$$[a, b]a = 0, \tag{3}$$

$$a(ba) - b(aa) = 0. \tag{4}$$

Let $\mathfrak{N}$ be a subvariety of the variety of right-commutative algebras defined by the identities (2), (3) and

$$a(ab) - 2a(ba) + b(aa) = 0. \tag{5}$$

In this work, we consider the free algebras $F_n(\mathfrak{M})$ and $F_n(\mathfrak{N})$ of rank n in the variety $\mathfrak{M}$ and $\mathfrak{N}$, respectively. We study the S_n -module structures of the free algebras $F_n(\mathfrak{M})$ and $F_n(\mathfrak{N})$, and show that the varieties $\mathfrak{M}$ and $\mathfrak{N}$ satisfy the distributive conditions. Moreover, we describe all subvarieties of the varieties $\mathfrak{M}$ and $\mathfrak{N}$.

II. METHODS

In 1950, Malcev [13] and Specht [14] were the first to independently apply the representation theory of the symmetric group to the study of polynomial identity algebras (PI-algebras). In the 1970s, Regev [15] further developed the Malcev–Specht method. In the 1980s, Drensky [16] and Berele [17] began applying the representation theory of the general linear group to the study of PI-algebras.

In 1986, Vladimirova and Drensky [18], using methods from the representation theory of the symmetric and general linear groups, completely described all subvarieties of the variety of associative algebras satisfying an identity of degree three over a field of characteristic zero. In the present paper, we apply the methods developed in [18].

Let $X_n = \{x_1, \dots, x_n\}$ be a set of generators, and let $K\{X_n\}$ be the absolutely free algebra on X_n . Let $\Delta_{x_i \mapsto x_j}$ be the derivation of $K\{X_n\}$ defined by

$$\begin{aligned}\Delta_{x_i \mapsto x_j}(x_i) &= x_j, \\ \Delta_{x_i \mapsto x_j}(x_k) &= 0, \quad k \neq i.\end{aligned}$$

III. RESULTS AND DISCUSSION

Throughout this paper, the field K is of characteristic zero. Let $P_n(\mathfrak{M})$ be a multilinear part of the free algebra $F_n(\mathfrak{M})$. Let $V(\lambda)$ be an irreducible S_n -module for partition $\lambda \vdash n$.

Proposition III.1. *As S_n -modules*

$$\begin{aligned}P_1(\mathfrak{M}) &\cong V(1), \quad P_2(\mathfrak{M}) \cong V(2) \oplus V(1, 1), \\ P_3(\mathfrak{M}) &\cong V(3) \oplus V(2, 1) \oplus V(1, 1, 1), \\ P_n(\mathfrak{M}) &\cong V(n) \oplus V(n-1, 1), \quad n \geq 4.\end{aligned}$$

The submodules $V(n)$ and $V(n-1, 1)$ are generated by the linearizations of $\underbrace{a(\cdots(aa)\cdots)}_n$ and $\underbrace{a(\cdots(a[a, b])\cdots)}_{n-2}$, respectively.

The submodule $V(1, 1, 1)$ is generated by $a(bc) - a(cb) + c(ab) - b(ac) + \underbrace{b(ca)}_n - \underbrace{c(ba)}_{n-2}$.

Proof. From (2) and (3) follows

$$(ab)c = 0.$$

From the last identity directly follow

$$((ab)c)d = 0, \tag{6}$$

$$(ab)(cd) = 0, \tag{7}$$

$$d((ab)c) = 0. \tag{8}$$

Applying (7) and (8) to the expression

$$\Delta_{a \mapsto aa} a(ba) - b(aa) = 0$$

we get

$$a(b(aa)) - b(a(aa)) = 0.$$

Multiplying (4) by a on the left, we get

$$a(a(ba)) - a(b(aa)) = 0.$$

Multiplying (4) by b on the left, we obtain

$$b(a(ba)) - b(b(aa)) = 0. \tag{9}$$

After swapping a and b in (9), we get

$$a(b(ab)) - a(a(bb)) = 0. \tag{10}$$

Applying (7) and (8) to the expression

$$\Delta_{a \mapsto ab} a(ba) - b(aa) = 0$$

we get

$$a(b(ab)) - b(a(ab)) = 0. \tag{11}$$

After swapping a and b in (11), we get

$$b(a(ba)) - a(b(ba)) = 0. \tag{12}$$

Multiplying (4) by c on the left, we obtain

$$c(a(ba)) - c(b(aa)) = 0. \quad (13)$$

Interchanging elements b and c in (13), we get

$$b(a(ca)) - b(c(aa)) = 0. \quad (14)$$

Using (7) and (8) in the expression

$$\Delta_{a \mapsto cd} a(ba) - b(aa) = 0$$

we get

$$a(b(cd)) - b(a(cd)) = 0. \quad (15)$$

From (13), (14) and (15) follow

$$a(b(ca)) = a(c(ba)) = b(a(ca)) = b(c(aa)) = c(a(ba)) = c(b(aa)). \quad (16)$$

Applying (16) to the expression

$$\Delta_{a \mapsto c} a(a(ba)) - a(b(aa)) = 0$$

we obtain

$$a(a(bc)) - a(b(ac)) = 0. \quad (17)$$

Using (15) in the expression

$$\Delta_{a \mapsto d} a(a(bc)) - a(b(ac)) = 0$$

we get

$$2a(d(bc)) - d(b(ac)) - a(b(dc)) = 0. \quad (18)$$

Interchanging elements a and b in (18), we get

$$2b(d(ac)) - d(a(bc)) - b(a(dc)) = 0. \quad (19)$$

Using (15) in the difference of (18) and (19), we obtain

$$a(d(bc)) - b(d(ac)) = 0. \quad (20)$$

As a result, from (15) and (20) follow

$$c(b(ad)) = c(a(bd)) = b(c(ad)) = b(a(cd)) = a(c(bd)) = a(b(cd)). \quad (21)$$

□

Let $P_n(\mathfrak{N})$ be a multilinear component of the free algebra $F_n(\mathfrak{N})$.

Proposition III.2. *As S_n -modules*

$$P_1(\mathfrak{N}) \cong V(1), \quad P_2(\mathfrak{N}) \cong V(2) \oplus V(1, 1),$$

$$P_3(\mathfrak{N}) \cong V(3) \oplus V(2, 1) \oplus V(1, 1, 1),$$

$$P_n(\mathfrak{N}) \cong V(0), \quad n \geq 4.$$

The submodules $V(1)$, $V(2)$ and $V(1, 1)$ are generated by the linearizations of a , aa and $ab - ba$, respectively.

The submodules $V(3)$, $V(2, 1)$ and $V(1, 1, 1)$ are generated by the linearizations of $a(aa)$, $a(ab) - a(ba)$ and $a(bc) - a(cb) + c(ab) - b(ac) + b(ca) - c(ba)$, respectively. .

Proof. From (2) and (3) follows

$$(ab)c = 0.$$

From the last identity directly follow

$$((ab)c)d = a((bc)d) = (ab)(cd) = 0. \quad (22)$$

Applying (22) to the expression $\Delta_{b \mapsto aa}(5)$, we get

$$a(a(aa)) = 0. \quad (23)$$

Applying (22) to the expression $\Delta_{b \mapsto ab}(5)$, we obtain

$$a(a(ab)) = 0. \quad (24)$$

Applying (22) to the expression $\Delta_{b \mapsto ba}(5)$, we obtain

$$a(a(ba)) = 0. \quad (25)$$

Applying (22) to the expression $\Delta_{a \mapsto aa}(5)$, we obtain

$$2a(b(aa)) - b(a(aa)) = 0. \quad (26)$$

Applying (24) and (25) to the expression $\Delta_{a \mapsto b}(23)$, we obtain

$$a(b(aa)) + b(a(aa)) = 0. \quad (27)$$

Summing of (26) and (27), we get

$$a(b(aa)) = 0. \quad (28)$$

Applying (28) to (27), we obtain

$$b(a(aa)) = 0. \quad (29)$$

Applying (22) to the expression $\Delta_{b \mapsto bb}(5)$, we get

$$a(a(bb)) = 0. \quad (30)$$

Applying (22) to the expression $\Delta_{a \mapsto ab}(5)$, we get

$$2a(b(ab)) - b(a(ab)) = 0. \quad (31)$$

Using (30) in the expression $\Delta_{a \mapsto b}(24)$, we get

$$a(b(ab)) + b(a(ab)) = 0. \quad (32)$$

Summing of (31) and (32), we obtain

$$a(b(ab)) = 0. \quad (33)$$

Applying (33) to (32), we get

$$b(a(ab)) = 0. \quad (34)$$

After swapping a and b in (30), (33) and (34), we get, respectively

$$b(b(aa)) = 0, \quad (35)$$

$$b(a(ba)) = 0, \quad (36)$$

$$a(b(ba)) = 0. \quad (37)$$

Applying (22) to the expression $\Delta_{b \mapsto bc}(5)$, we get

$$a(a(bc)) = 0. \quad (38)$$

Applying (22) to the expression $\Delta_{a \mapsto ac}(5)$, we get

$$2a(b(ac)) - b(a(ac)) = 0. \quad (39)$$

Applying (38) to the expression $\Delta_{a \mapsto b} a(a(c)) = 0$, we obtain

$$b(a(ac)) + a(b(ac)) = 0. \quad (40)$$

Summing of (39) and (40), we get

$$a(b(ac)) = 0. \quad (41)$$

Applying (41) to (40), we get

$$b(a(ac)) = 0. \quad (42)$$

Using (22) in the expression $\Delta_{a \mapsto ca}(5)$, we get

$$2a(b(ca)) - b(a(ca)) = 0. \quad (43)$$

Applying (38) to the expression $\Delta_{a \mapsto b} a(ca) = 0$, we obtain

$$b(a(ca)) + a(b(ca)) = 0. \quad (44)$$

Summing of (43) and (44), we get

$$a(b(ca)) = 0. \quad (45)$$

Applying (45) to (44), we obtain

$$b(a(ca)) = 0. \quad (46)$$

Applying (46) and (42) to the expression $\Delta_{a \mapsto c} b(aa) = 0$, we get

$$b(c(aa)) = 0. \quad (47)$$

Applying (22) to $\Delta_{a \mapsto cd}(5)$, we get

$$2a(b(cd)) - b(a(cd)) = 0. \quad (48)$$

Using the derivation $\Delta_{a \mapsto b}$ in the expression $a(cd) = 0$, we obtain

$$a(b(cd)) + b(a(cd)) = 0. \quad (49)$$

Summing of (48) and (49), we obtain

$$a(b(cd)) = 0.$$

□

Theorem III.1. Let $f \in F_n(\mathfrak{M})$. Then the consequences of higher degree of f are equivalent to the following identities

- a) $a(bc) - a(cb) + c(ab) - b(ac) + b(ca) - c(ba)$, if $f = aa$;
- b) $a(bc) - a(cb) + c(ab) - b(ac) + b(ca) - c(ba)$, if $f = ab - ba$;
- c) $\underbrace{a(\cdots(a(aa))\cdots)}_{n+1}$ and $\underbrace{a(\cdots(a[a, b])\cdots)}_{n-1}$, if $f = \underbrace{a(\cdots(a(aa))\cdots)}_n$, $n \geq 1$;
- d) $\underbrace{a(\cdots(a(aa))\cdots)}_{n+1}$ and $\underbrace{a(\cdots(a[a, b])\cdots)}_{n-1}$, if $f = \underbrace{a(\cdots(a[a, b])\cdots)}_{n-2}$, $n \geq 2$.

Proof. a) Applying $aa = 0$ to (4), we get

$$(ba)a = 0.$$

Using the derivation $\Delta_{a \mapsto c}$ in the identity $(ba)a = 0$, we obtain

$$(bc)a + (ba)c = 0.$$

Applying (1) to the identity $(bc)a + (ba)c = 0$, we get

$$(bc)a = 0.$$

Since $a(bc) = -(bc)a$. Hence,

$$a(bc) - a(cb) + c(ab) - b(ac) + b(ca) - c(ba) = 0.$$

b) It is clear.

$$c) \underbrace{a(\cdots(aa)\cdots)}_{n+1} = a(\underbrace{a(\cdots(aa)\cdots)}_n) = 0.$$

Let $n = 1$. It is clear.

Let $n = 2$. Then from $aa = 0$ follows $b(aa) = 0$. From identities $b(aa) = 0$ and (4) follows $a(ba) = 0$. Applying identities $b(aa) = 0$ and $a(ba) = 0$ to the expression

$$\Delta_{a \mapsto b} a(aa) = 0$$

we obtain

$$a(ab) = 0.$$

Hence, $a[a, b] = a(ab) - a(ba) = 0$.

Let $n > 2$. Applying identities (21) and $b(\underbrace{a(\cdots(aa)\cdots)}_n) = 0$ to the expression

$$\Delta_{a \mapsto b} \underbrace{a(\cdots(aa)\cdots)}_{n+1} = 0$$

we obtain

$$a(\cdots(ab)\cdots) = 0$$

Hence, $\underbrace{a(\cdots(a[a, b])\cdots)}_{n-1} = 0$.

$$d) \underbrace{a(\cdots(a[a, b])\cdots)}_{n-1} = a(\underbrace{a(\cdots(a[a, b])\cdots)}_{n-2}) = 0.$$

Applying identities (2) and (21) to the expression

$$\Delta_{b \mapsto aa} \underbrace{a(\cdots(a[a, b])\cdots)}_{n-2} = 0$$

we get

$$\underbrace{a(\cdots(aa)\cdots)}_{n+1} = 0.$$

□

Theorem III.2. Let $f \in F_n(\mathfrak{N})$. Then the consequences of higher degree of f are equivalent to the following identities

$$a) aa, ab - ba, \text{ if } f = a;$$

$$b) a(aa), a[a, b], a(bc) - a(cb) + c(ab) - b(ac) + b(ca) - c(ba), \text{ if } f = aa;$$

$$c) a[a, b], a(bc) - a(cb) + c(ab) - b(ac) + b(ca) - c(ba), \text{ if } f = ab - ba.$$

Proof. a) It is clear.

b) From (2) and (3) follows $(ab)c = 0$. Applying the last identity to the expression $\Delta_{a \mapsto bc} aa = 0$, we obtain $a(bc) = 0$.

c) It is clear.

□

IV. CONCLUSION AND FUTURE WORK

In this work, we studied the free algebras in the varieties $\mathfrak{M}$ and $\mathfrak{N}$. We investigated the S_n -module structures of the free algebras $F_n(\mathfrak{M})$ and $F_n(\mathfrak{N})$. As a consequence, we proved that the varieties $\mathfrak{M}$ and $\mathfrak{N}$ satisfy the distributivity conditions. Moreover, we described all subvarieties of the varieties $\mathfrak{M}$ and $\mathfrak{N}$.

It would be of interest to investigate whether there exist further subclasses of the class of right-commutative algebras satisfying the distributivity conditions. Moreover, it would be interesting to study operads encoding the above mentioned subvarieties of the variety of right-commutative algebras.

V. ACKNOWLEDGMENT

The authors would like to express their gratitude to SDU University for providing the necessary resources and facilities to conduct this study.

REFERENCES

- [1] M. Livernet, A rigidity theorem for pre-Lie algebras, *Journal of Pure and Applied Algebra*, Volume 207, Issue 1, September 2006, Pages 1-18
- [2] G.W. Zinbiel, *Encyclopedia of types of algebras 2010*, Nankai Series in Pure, Applied Mathematics and Theoretical Physics, Operads and Universal Algebra, pp. 217-297 (2012)
- [3] A.S. Dzhumadil'daev, Codimension Growth and Non-Koszulity of Novikov Operad, *Communications in Algebra*, V. 39 (2011), 2943–2952
- [4] A.S. Dzhumadil'daev, N.A. Ismailov, S_n - and GL_n -module structures on free Novikov algebras, *Journal of Algebra*, Volume 416, 15 October 2014, Pages 287-313
- [5] A.S. Dzhumadil'daev, K.M. Tulenbaev, Bicommutative algebras, *Russian Mathematical Surveys*, Volume 58, Issue 6, (2003), 1196–1197
- [6] A.S. Dzhumadil'daev, N.A. Ismailov, K.M. Tulenbaev, Free bicommutative algebras, *Serdica Math. J.* V. 37 (2011), 25–44
- [7] V.S. Drensky, B.K. Zhakhayev, Noetherianity and Specht problem for varieties of bicommutative algebras, *Journal of Algebra*, V. 499 (2018), 570–582
- [8] V.T. Filippov, V.K. Kharchenko, I.P. Shestakov (Eds.), *Dniester Notebook: Unsolved Problems in the Theory of Rings and Modules* (in Russian), Institute of Mathematics, Siberian Branch of Academy of Sciences USSR, Novosibirsk, 1976
- [9] A.Z. Ananin, A.R. Kemer, Varieties of associative algebras whose lattices of subvarieties are distributive. In: *Sibirsk. Mat. Ž.* 17.4 (1976), pp. 723–730
- [10] V.D. Martirosyan, On the distributivity of lattices of subvarieties of varieties of alternative algebras. In: *Mat. Sb. (N.S.)* 118(160).1 (1982), pp. 118–137, 143
- [11] V.D. Martirosyan, Distributivity of lattices of subvarieties of varieties of right-alternative algebras. In: *Akad. Nauk Armyan. SSR Dokl.* 78.5(1984), pp. 199–202
- [12] V. Dotsenko, B.K. Zhakhayev, Distributive lattices of varieties of Novikov algebras, *Manuscripta Mathematica*, Volume 176, article number 29, (2025)
- [13] A.I. Malcev, On algebras defined by identities // *Mat. Sbornik N.S.* -1950. -Vol. 26 (68), No. 1. -P. 19-33
- [14] W. Specht, Gesetze in Ringen I // *Math. Z.* -V. 52 (1950). -P. 557-589
- [15] Regev A. On the codimensions of matrix algebras // *Lecture Notes in Math.* -1988. -Vol. 1352. -P. 162-172
- [16] Drensky V. Representations of the symmetric group and varieties of linear algebras (Russian) // *Mat. Sb.* -1981. -Vol. 115. -P. 98-115. Translation: *Math. USSR Sb.* -1981. -Vol. 43. -P. 85-101
- [17] Berele A. Homogeneous polynomial identities // *Israel J. Math.* -1982. -Vol. 42. -P. 258-272
- [18] L.A. Vladimirova, V.S. Drensky, Varieties of associative algebras with third-degree identity // *Pliska Studia Mathematica Bulgarica* -1986. -V. 8. -P. 144-157 (in Russian)

Article

Space of Tortken elements in free Novikov algebra

Baigali Meruyert ^{1*}

¹School of Information Technology and Applied Mathematics, SDU University, Almaty, Kazakhstan, SDU University, Kaskelen, Kazakhstan

DOI: 10.47344/0wq3j174

Abstract

In this paper, we study the one-generated case of the space of Tortken elements in a free Novikov algebra. Using the differential-monomial realization, we prove that the dimensions of the homogeneous components are given by $\dim T_1 = \dim T_2 = 1$ and $\dim T_n = p(n-2)$ for all $n \geq 3$, where $p(n-2)$ denotes the partition number. We also describe the first homogeneous components in the circle language and analyze the relations among their elements in low degrees. In particular, we show that the elements are linearly independent up to degree 5, while the first nontrivial relation appears in degree 6. We derive explicit identities in degrees 6 and 7 and prove a recursive inclusion for homogeneous components in the one-generated case for $n \leq 7$. These results give a more precise description of the low-degree structure of the space of Tortken elements and suggest a regular pattern that may persist in higher degrees. We also formulate a conjecture extending the recursive inclusion to all degrees.

Keywords: Novikov algebra, Tortken algebra, symmetric elements, free algebra

I. INTRODUCTION

An algebra $(A, \cdot)$ is called a (right) Novikov algebra if, for all $a, b, c \in A$, the following identities hold:

$$a \cdot (b \cdot c) = b \cdot (a \cdot c), \quad (1)$$

$$(a \cdot b) \cdot c - a \cdot (b \cdot c) = (a \cdot c) \cdot b - a \cdot (c \cdot b). \quad (2)$$

The first identity (1) is called the **left-commutativity** identity, and the second identity (2) is called the **right-symmetric** identity, and any algebra satisfying this identity is called the right-symmetric algebra. Thus, throughout the paper we use the convention that a (right) Novikov algebra is an algebra satisfying identities (1) and (2).

Example 1.1. Let us consider the polynomial algebra $A = K[x]$ with multiplication

$$a \cdot b = D(a)b$$

A direct verification shows that $(A, \cdot)$ is a Novikov algebra.

*Corresponding author: 241101007@sdu.edu.kz

Email: 241101007@sdu.edu.kz ORCID: 0009-0004-5443-1559

Received: April 15, 2026. Reviewed: May 14, 2026. Accepted: September 30, 2026. © 2026 Baigali Meruyert. All rights reserved.

Now, let us define a new product on the Novikov algebra, which is called a symmetric product (anti-commutator). Let A be a (right) Novikov algebra with multiplication $a \cdot b$. Then the anti-commutator is defined by:

$$a \circ b := a \cdot b + b \cdot a$$

Notice that $a \circ b = b \circ a$.

Dzhumadil'daev [1], [3] proved that the symmetrized product satisfies the following identities:

$$(a \circ b) \circ (c \circ d) - (a \circ d) \circ (c \circ b) = (a, b, c)^+ \circ d - (a, d, c)^+ \circ b, \quad (3)$$

$$\begin{aligned} & (((a \circ a) \circ a) \circ b) \circ b + (((a \circ b) \circ b) \circ a) \circ a \\ & + 2(((a \circ a) \circ b) \circ b) \circ a + 2(((a \circ b) \circ a) \circ a) \circ b \\ & - 3(((a \circ a) \circ b) \circ a) \circ b - 3(((a \circ b) \circ a) \circ b) \circ a = 0 \end{aligned} \quad (4)$$

where

$$(a, b, c)^+ = a \circ (b \circ c) - (a \circ b) \circ c.$$

The first identity is called the Tortken identity, and the second identity is the Besken identity. Algebras that satisfy (3) are called Tortken algebras.

Let us define the free Novikov algebra generated by a set X and call it $Nov\langle X \rangle$. Then the space of symmetric elements $S\langle X \rangle$ is the smallest subspace of $Nov\langle X \rangle$ that contains X and is closed under the $\circ$ -product. Elements of $S\langle X \rangle$ are called symmetric elements or Tortken elements, otherwise.

Now, let us define $\mathcal{M}$ as a set of all differential monomials with one derivation. A basis of the free Novikov algebras was constructed by Dzhumadil'daev and Löfwall [2]:

Theorem I.2. *The set*

$$\mathcal{N} = \{ u \in \mathcal{M} \mid \deg(u) - d(u) = 1 \}$$

is a basis of the free Novikov algebra $Nov\langle X \rangle$.

And recently, Dzhumadil'daev and Ismailov [4] constructed a basis of the space of Tortken elements in the free Novikov algebra in terms of differential polynomials:

Theorem I.3. *The set*

$$\overline{\mathcal{M}} := X \cup \{ u' \mid u \in \mathcal{M}, \deg(u) - d(u) = 2 \}$$

is a basis of the space of Tortken elements $S\langle X \rangle$.

In this paper, we investigate the space of Tortken elements in a free Novikov algebra in the one-generated case. While a basis in terms of differential monomials is already known [4], the corresponding description in terms of $\circ$ -product remains unclear. Our aim is to describe this basis in low degrees and to compare it with the known differential-monomial model. We show that the dimensions of the homogeneous components are governed by the partition numbers $p(n-2)$, and we construct explicit basis elements in the $\circ$ -language for small degrees. The necessary identities are verified up to degree seven by means of computations carried out in *Wolfram Mathematica*. This work may be regarded as a step toward the study of free special Tortken algebras in the $\circ$ -product language.

II. METHODS

Proposition II.1. *Let T_n be the homogeneous component of degree n of the space of Tortken elements in the free Novikov algebra generated by one element x . Then*

$$\dim T_1 = \dim T_2 = 1, \quad \dim T_n = p(n-2) \quad \text{for all } n \geq 3,$$

where $p(m)$ denotes the number of partitions of m .

Proof. By Theorem I.3, a basis of $S\langle x \rangle$ is given by

$$\overline{\mathcal{M}} := X \cup \{ u' \mid u \in \mathcal{M}, \deg(u) - d(u) = 2 \},$$

where $\mathcal{M}$ is the set of differential monomials. In the one-generated case, every differential monomial can be written uniquely in the form

$$u = x^{(i_1)} x^{(i_2)} \dots x^{(i_n)}$$

Moreover,

$$\deg(u) = n, \quad d(u) = i_1 + \dots + i_n.$$

Hence the condition $\deg(u) - d(u) = 2$ becomes

$$i_1 + \dots + i_n = n - 2.$$

Since all factors involve the same generator x and the product is commutative, we may assume without loss of generality that the derivation orders are arranged in nonincreasing order:

$$i_1 \geq i_2 \geq \dots \geq i_n \geq 0.$$

Then the number such nonincreasing sequences is equal to the number of partitions of $n - 2$. Therefore,

$$\dim T_n = p(n - 2), \quad n \geq 2.$$

Finally, for $n = 1$, the homogeneous component T_1 is spanned by x . For $n = 2$, the component T_2 is spanned by $x \circ x$. Hence,

$$\dim T_1 = \dim T_2 = 1.$$

This proves the claim. □

Proposition II.2. *Let T_n be the homogeneous component of degree n in the one-generated case. Then,*

$$T_n \subseteq T_{n-1} \circ T_1 + T_{n-2} \circ T_2, \quad n \leq 7.$$

Proof. In the one-generator case, bases in the $\circ$ -language for the first degrees may be chosen as follows:

$$T_1 = \langle x \rangle, \quad T_2 = \langle x \circ x \rangle, \quad T_3 = \langle (x \circ x) \circ x \rangle,$$

The first three degrees are immediate, since each homogeneous component is one-dimensional.

A. Degree $n = 4$

Degree 4 is the first case in which the homogeneous component has more than one basis element, so we write it out in detail. Let

$$w_1 = ((x \circ x) \circ x) \circ x, \quad w_2 = (x \circ x) \circ (x \circ x).$$

The differential basis of T_4 is given by

$$b_1 = (x''xxx)' = x'''xxx + 3x''x'xx, \quad b_2 = (x'x'xx)' = 2x''x'xx + 2x'x'x'x.$$

Using the identity $a \circ b = (ab)'$, we obtain

$$w_1 = ((x \circ x) \circ x) \circ x = 2x'''xxx + 14x''x'xx + 8x'x'x'x$$

$$w_2 = (x \circ x) \circ (x \circ x) = 8x''x'xx + 8x'x'x'x$$

$$2x'''xxx + 14x''x'xx + 8x'x'x'x = \alpha_1 b_1 + \alpha_2 b_2 \Rightarrow \alpha_1 = 2, \quad \alpha_2 = 4$$

$$8x''x'xx + 8x'x'x'x = \alpha_1 b_1 + \alpha_2 b_2 \Rightarrow \alpha_1 = 0, \quad \alpha_2 = 4$$

That is,

$$w_1 = 2b_1 + 4b_2, \quad w_2 = 4b_2.$$

w_1 and w_2 are linearly independent.

B. Degree $n = 5$

Since

$$\dim T_5 = p(3) = 3$$

in degree $n = 5$, we have three basis elements which are

$$w_1 = (((x \circ x) \circ x) \circ x) \circ x, \quad w_2 = ((x \circ x) \circ (x \circ x)) \circ x, \quad w_3 = ((x \circ x) \circ x) \circ (x \circ x).$$

The differential basis of T_5 is given by

$$b_1 = (x'''xxxx)', \quad b_2 = (x''x'xxx)', \quad b_3 = (x'x'x'xx)'$$

Therefore, we compute

$$w_1 = 2b_1 + 14b_2 + 8b_3, \quad w_2 = 8b_2 + 8b_3, \quad w_3 = 4b_2 + 8b_3$$

Here, w_1, w_2 and w_3 are linearly independent.

C. Degree $n = 6$

The first degree where linear dependence among natural $\circ$ -words appears is degree 6. Consider the six natural commutative nonassociative words

$$\begin{aligned} w_1 &= (((x \circ x) \circ x) \circ x) \circ x, \\ w_2 &= (((x \circ x) \circ (x \circ x)) \circ x) \circ x, \\ w_3 &= (((x \circ x) \circ x) \circ (x \circ x)) \circ x, \\ w_4 &= (((x \circ x) \circ x) \circ x) \circ (x \circ x), \\ w_5 &= ((x \circ x) \circ (x \circ x)) \circ (x \circ x), \\ w_6 &= (x \circ (x \circ x)) \circ (x \circ (x \circ x)). \end{aligned}$$

Since

$$\dim T_6 = p(4) = 5,$$

these six elements cannot be linearly independent.

To construct the identity in degree 6, we use *Wolfram Mathematica*. As a first step, we implement the derivation operator and the symmetrized product in the commutative differential algebra by encoding their basic properties; see Appendix A.

Then we enumerate all six elements of degree 6 in the $\circ$ -language and list all seven basis elements of the homogeneous component of degree 6 of the free Novikov algebra on one generator and assign to each element w_i a scalar coefficient λ_i and consider the linear combination

$$\sum_{i=1}^6 \lambda_i w_i = 0.$$

The program then determines the relations among the coefficients λ_i and computes their admissible values; see Appendix B.

The resulting relations among the coefficients in the case $n = 6$ are

$$\begin{aligned} \lambda_1 &= 0, \\ \lambda_3 &= -\frac{\lambda_2}{2}, \\ \lambda_4 &= -\frac{\lambda_2}{2}, \\ \lambda_5 &= \frac{\lambda_2}{2}, \\ \lambda_6 &= -\frac{\lambda_2}{2}. \end{aligned}$$

Taking $\lambda_2 = 2$, we obtain the following identity:

$$\begin{aligned} & 2((x \circ x) \circ (x \circ x)) \circ (x \circ x) - ((x \circ x) \circ x) \circ ((x \circ x) \circ x) \\ & - (((x \circ x) \circ x) \circ x) \circ (x \circ x) + (((x \circ x) \circ (x \circ x)) \circ x) \circ x \\ & - (((x \circ x) \circ x) \circ (x \circ x)) \circ x = 0 \end{aligned} \quad (5)$$

or equivalently,

$$2w_5 - w_6 - w_4 + w_2 - w_3 = 0.$$

This identity is a consequence of the Tortken identity with

$$a = d = x \circ x, \quad b = c = x.$$

Substituting these values into the Tortken identity gives

$$((x \circ x) \circ x) \circ (x \circ (x \circ x)) - ((x \circ x) \circ (x \circ x)) \circ (x \circ x) = (x \circ x, x, x)^+ \circ (x \circ x) - (x \circ x, x \circ x, x)^+ \circ x.$$

Indeed, this substitution gives

$$-2w_5 + w_6 + w_4 - w_2 + w_3 = 0,$$

or equivalently,

$$2w_5 - w_6 - w_4 + w_2 - w_3 = 0.$$

Therefore, we can eliminate the element w_6 :

$$w_6 = w_2 - w_3 - w_4 + 2w_5.$$

Notice that

$$w_6 \in T_3 \circ T_3$$

and

$$w_1 \in (((T_2 \circ T_1) \circ T_1) \circ T_1) \circ T_1$$

where $((T_2 \circ T_1) \circ T_1) \circ T_1 \in T_5$. That is,

$$w_1 \in T_5 \circ T_1.$$

The same holds for w_2 and w_3 ,

$$w_2 \in T_5 \circ T_1, \quad w_3 \in T_5 \circ T_1.$$

And for w_4 and w_5 we define

$$w_4 \in T_4 \circ T_2, \quad w_5 \in T_4 \circ T_2.$$

From the identity $w_6 = w_2 - w_3 - w_4 + 2w_5$, it follows that

$$T_3 \circ T_3 \subseteq T_5 \circ T_1 + T_4 \circ T_2.$$

That is,

$$T_6 \subseteq T_5 \circ T_1 + T_4 \circ T_2.$$

D. Degree $n = 7$

For degree $n = 7$, we have 11 elements:

$$\begin{aligned} w_1 &= (((((x \circ x) \circ x) \circ x) \circ x) \circ x) \circ x, & w_2 &= (((((x \circ x) \circ (x \circ x)) \circ x) \circ x) \circ x) \circ x, \\ w_3 &= (((((x \circ x) \circ x) \circ (x \circ x)) \circ x) \circ x) \circ x, & w_4 &= (((((x \circ x) \circ x) \circ x) \circ (x \circ x)) \circ x) \circ x, \\ w_5 &= (((((x \circ x) \circ (x \circ x)) \circ (x \circ x)) \circ x) \circ x) \circ x, & w_6 &= (((((x \circ x) \circ x) \circ (x \circ x) \circ x)) \circ x) \circ x, \\ w_7 &= (((((x \circ x) \circ x) \circ x) \circ x) \circ (x \circ x)) \circ x) \circ x, & w_8 &= (((((x \circ x) \circ (x \circ x)) \circ x) \circ (x \circ x)) \circ x) \circ x, \\ w_9 &= (((((x \circ x) \circ x) \circ (x \circ x)) \circ (x \circ x)) \circ x) \circ x, & w_{10} &= (((((x \circ x) \circ x) \circ x) \circ (x \circ x) \circ x)) \circ x, \\ w_{11} &= (((((x \circ x) \circ (x \circ x)) \circ (x \circ x) \circ x)) \circ x) \circ x. \end{aligned}$$

Since $p(7-2) = p(5) = 7$, these 11 elements cannot be linearly independent. Thus, four linear relations occur among these eleven candidate elements.: We eliminate w_6 since in the $n = 6$ case we concluded that

$$T_3 \circ T_3 \subseteq T_5 \circ T_1 + T_4 \circ T_2.$$

For our calculations, we will use *Wolfram Mathematica* (see Appendix C). By this, we obtained

$$\begin{aligned} \lambda_1 &= 0, \\ \lambda_2 &= \lambda_{10} + 2\lambda_{11}, \\ \lambda_4 &= -7\lambda_{10} - 12\lambda_{11} - 3\lambda_3, \\ \lambda_5 &= 6\lambda_{10} + 10\lambda_{11} + 2\lambda_3, \\ \lambda_7 &= 5\lambda_{10} + 8\lambda_{11} + 2\lambda_3, \\ \lambda_8 &= -6\lambda_{10} - 9\lambda_{11} - 3\lambda_3, \\ \lambda_9 &= \lambda_3. \end{aligned}$$

Taking

$$(\lambda_3, \lambda_{10}, \lambda_{11}) = (0, 1, 0), \quad (0, 0, 1), \quad (1, 0, 0),$$

we obtain the following three identities:

$$w_{10} + w_2 - 7w_4 + 6w_5 + 5w_7 - 6w_8 = 0, \quad (6a)$$

$$w_{11} + 2w_2 - 12w_4 + 10w_5 + 8w_7 - 9w_8 = 0, \quad (6b)$$

$$w_3 - 3w_4 + 2w_5 + 2w_7 - 3w_8 + w_9 = 0. \quad (6c)$$

The third relation (6c) coincides with the jetken polynomial, which is obtained from the Besken identity by taking $a = x$ and $b = x \circ x$ [1]. Identity (6a) allows us to eliminate w_{10} , while identity (6b) allows us to eliminate w_{11} . The third identity (6c) eliminates one more element from the remaining set. Thus, we obtain 7 linearly independent elements.

Since

$$w_{10}, w_{11} \in T_4 \circ T_3, \quad w_2, w_4, w_5 \in T_6 \circ T_1, \quad w_7, w_8 \in T_5 \circ T_2,$$

and

$$w_{10} = -w_2 + 7w_4 - 6w_5 - 5w_7 + 6w_8, \quad w_{11} = -2w_2 + 12w_4 - 10w_5 - 8w_7 + 9w_8,$$

it follows that

$$T_4 \circ T_3 \subseteq T_6 \circ T_1 + T_5 \circ T_2.$$

That is,

$$T_7 \subseteq T_6 \circ T_1 + T_5 \circ T_2.$$

□

These computations motivate the following conjecture below. Our computations show that, up to degree 7, the stronger recursive inclusion holds.

Conjecture II.3. Let T_n be the homogeneous component of degree n in the one-generated case. Then,

$$T_n \subseteq T_{n-1} \circ T_1 + T_{n-2} \circ T_2, \quad n \geq 3.$$

These observations may be helpful in the study of the problem at higher degrees.

III. RESULTS AND DISCUSSION

In this work, we studied the homogeneous components T_n in the one-generated case. Our first main result gives the exact dimensions of these components:

$$\dim T_1 = \dim T_2 = 1, \quad \dim T_n = p(n-2) \quad \text{for all } n \geq 3.$$

Thus, the dimensions of the homogeneous components are described in combinatorial terms by the partition numbers.

We next describe the first homogeneous components in the $\circ$ -language and analyze the relations among their candidate basis elements. In degrees 1 through 5, the elements under consideration are linearly independent. The first nontrivial relation appears in degree 6. More precisely, by passing from the $\circ$ -language to the differential-monomial realization of the free Novikov algebra, we obtain identity (5), which is a consequence of the Tortken identity. This shows that in degree 6 the list of candidate elements is no longer independent, and one of them can be expressed through the others.

The same persists in degree 7, where further relations arise among the candidate elements. Together with the identity obtained in degree 6, these relations refine the structure of the homogeneous component T_7 . Thus, beginning with degree 6, the description of T_n requires not only the enumeration of all possible $\circ$ -elements, but also a careful analysis of the identities they satisfy.

Another main result is the recursive inclusion

$$T_n \subseteq T_{n-1} \circ T_1 + T_{n-2} \circ T_2, \quad n \leq 7.$$

This inclusion gives a relation between consecutive homogeneous components and shows that, in low degrees, elements of T_n can be obtained from elements of smaller degree by means of the symmetrized product.

A. Discussion

The obtained results show that even the one-generated case is already nontrivial. Although the dimensions of the homogeneous components admit a simple description, the identities appearing in degrees 6 and 7 show that the space of symmetric elements cannot be understood by listing $\circ$ -elements alone. One must also take into account the relations between them.

These observations indicate that the one-generated case already reflects essential features of Tortken elements. In particular, the low-degree computations reveal a regular pattern in the structure of the homogeneous components, which may remain visible in higher degrees.

The results obtained here may therefore be useful for the further study of Tortken elements in higher degrees and for the investigation of the case of several generators.

IV. CONCLUSION AND FUTURE WORK

In this paper, we studied the one-generated case of the space of Tortken elements in a free Novikov algebra. We proved the dimension formula

$$\dim T_1 = \dim T_2 = 1, \quad \dim T_n = p(n-2) \quad \text{for all } n \geq 3,$$

described the first homogeneous components in the $\circ$ -language, obtained explicit identities in degrees 6 and 7, and established the recursive inclusion

$$T_n \subseteq T_{n-1} \circ T_1 + T_{n-2} \circ T_2, \quad n \leq 7.$$

These results provide a precise description of the low-degree structure of the one-generated case. In particular, they indicate that the one-generated case already captures essential features of the structure of special Tortken algebras.

A natural problem for future work is to determine whether the recursive inclusion extends to all degrees. This leads to the conjecture

$$T_n \subseteq T_{n-1} \circ T_1 + T_{n-2} \circ T_2, \quad n \geq 3.$$

If true, this conjecture would provide a systematic way to describe the higher-degree components in the one-generated case. It may also serve as a useful starting point for studying the operadic and structural properties of special Tortken algebras and their connections with related Novikov-type algebras, such as δ -Novikov algebras, Gelfand–Dorfman algebras, and Novikov–Poisson algebras.

V. ACKNOWLEDGMENT

APPENDIX

A. Implementation of the derivation and the symmetrized product

The derivation was implemented by the following rules. The symmetrized product $a \circ b = (ab)'$ was implemented as $t[a, b]$.

```
Dr[C_Integer a_] := C Dr[a]

Dr[a_ + b_] := Dr[a] + Dr[b]

Dr[C_Integer a_ + K_Integer b_] := C Dr[a] + K Dr[b]

Dr[C_Integer a_ + K_Integer b_ + D_Integer d_] :=
  C Dr[a] + K Dr[b] + D Dr[d]

Dr[a_ b_] := Dr[a] b + a Dr[b]

Dr[a_^n_Integer] := Dr[a^(n - 1)] a + a^(n - 1) Dr[a]

Dr'[x_] := 0

Dr[0] := 0

t[a_, b_] := ExpandAll[Dr[a b]]
```

B. Degree 6 computation

Here we present the computation in degree 6. The following six candidate elements of degree six in the $\circ$ -language were used.

```
Genldeg6 = {
  gld6[1] = t[t[t[t[x, x], x], x], x], x],
  gld6[2] = t[t[t[x, x], t[x, x]], t[x, x]],
  gld6[3] = t[t[t[x, x], x], t[t[x, x], x]],
  gld6[4] = t[t[t[t[x, x], x], x], t[x, x]],
  gld6[5] = t[t[t[t[x, x], t[x, x]], x], x],
  gld6[6] = t[t[t[t[x, x], x], t[x, x]], x]
};
```

The program then determines the relations among the coefficients λ_i and computes their admissible values.

```
Combinldeg6 =
  Coefficient[
```

```

Collect[
  ExpandAll[Sum[Subscript[lambda, i] gld6[i], {i, 1, 6}]],
  {
    x Dr[x]^5,
    x^2 Dr[x]^3 Dr[Dr[x]],
    x^3 Dr[x] Dr[Dr[x]]^2,
    x^3 Dr[x]^2 Dr[Dr[Dr[x]]],
    x^4 Dr[Dr[x]]*Dr[Dr[Dr[x]]],
    x^4 Dr[x]*Dr[Dr[Dr[Dr[x]]]],
    x^5 Dr[Dr[Dr[Dr[Dr[x]]]]
  }
],
{
  x Dr[x]^5,
  x^2 Dr[x]^3 Dr[Dr[x]],
  x^3 Dr[x] Dr[Dr[x]]^2,
  x^3 Dr[x]^2 Dr[Dr[Dr[x]]],
  x^4 Dr[Dr[x]]*Dr[Dr[Dr[x]]],
  x^4 Dr[x]*Dr[Dr[Dr[Dr[x]]]],
  x^5 Dr[Dr[Dr[Dr[Dr[x]]]]
}
];

S[x_] := x == 0;

TableForm[
  First[
    Solve[
      Map[S, Combin1deg6],
      Array[Subscript[lambda, #] &, 6]
    ]
  ]
]

```

C. Degree 7 computation

Here we present the computation in degree 7.

```

Gen1deg7 = {
  gld7[1] = t[t[t[x, x], x], x], t[t[x, x], x]],
  gld7[2] = t[t[t[x, x], t[x, x]], t[t[x, x], x]],
  gld7[3] = t[t[t[t[t[x, x], x], x], x], x], x],
  gld7[4] = t[t[t[t[x, x], t[x, x]], x], x], x],
  gld7[5] = t[t[t[t[x, x], x], t[x, x]], x], x],
  gld7[6] = t[t[t[t[x, x], x], x], t[x, x]], x],
  gld7[7] = t[t[t[x, x], t[x, x]], t[x, x]], x],
  gld7[8] = t[t[t[t[x, x], x], x], x], t[x, x]],
  gld7[9] = t[t[t[x, x], t[x, x]], x], t[x, x]],
  gld7[10] = t[t[t[x, x], x], t[x, x]], t[x, x]]
};

Combin1deg7 =

```

```

Coefficient[
Collect[
ExpandAll[Sum[Subscript[\[Lambda], i] g1d7[i], {i, 1, 10}]],
{
x Dr[x]^6,
x^2 Dr[x]^4 Dr[Dr[x]],
x^3 Dr[x]^2 Dr[Dr[x]]^2,
x^4 Dr[Dr[x]]^3,
x^3 Dr[x]^3 Dr[Dr[Dr[x]]],
x^4 Dr[x] Dr[Dr[x]] Dr[Dr[Dr[x]]],
x^5 Dr[Dr[Dr[x]]]^2,
x^4 Dr[x]^2 Dr[Dr[Dr[Dr[x]]]],
x^5 Dr[Dr[x]] Dr[Dr[Dr[Dr[x]]]],
x^5 Dr[x] Dr[Dr[Dr[Dr[Dr[x]]]]],
x^6 Dr[Dr[Dr[Dr[Dr[Dr[x]]]]]]
}
],
{
x Dr[x]^6,
x^2 Dr[x]^4 Dr[Dr[x]],
x^3 Dr[x]^2 Dr[Dr[x]]^2,
x^4 Dr[Dr[x]]^3,
x^3 Dr[x]^3 Dr[Dr[Dr[x]]],
x^4 Dr[x] Dr[Dr[x]] Dr[Dr[Dr[x]]],
x^5 Dr[Dr[Dr[x]]]^2,
x^4 Dr[x]^2 Dr[Dr[Dr[Dr[x]]]],
x^5 Dr[Dr[x]] Dr[Dr[Dr[Dr[x]]]],
x^5 Dr[x] Dr[Dr[Dr[Dr[Dr[x]]]]],
x^6 Dr[Dr[Dr[Dr[Dr[Dr[x]]]]]]
}
];

S[x_] := x == 0;

TableForm[
First[
Solve[
Map[S, Combin1deg7],
Array[Subscript[\[Lambda], #] &, 10]
]
]
]

```

REFERENCES

- [1] A. S. Dzhumadil'daev, "Special identity for Novikov–Jordan algebras," *Communications in Algebra*, vol. 33, no. 5, pp. 1279–1287, 2005. DOI: 10.1081/AGB-200060504.
- [2] A. S. Dzhumadil'daev and C. Löfwall, "Trees, free right-symmetric algebras, free Novikov algebras and identities," *Homology, Homotopy and Applications*, vol. 4, no. 2, pp. 165–190, 2002. DOI: 10.4310/HHA.2002.v4.n2.a8.
- [3] A. S. Dzhumadil'daev, "Novikov–Jordan algebras," *Communications in Algebra*, vol. 30, no. 11, pp. 5207–5240, 2002. DOI: 10.1081/AGB-120015649.

- [4] A. S. Dzhumadil'daev and N. Ismailov, "Null Lagrangians in free Novikov algebras," arXiv preprint arXiv:2601.11168, 2026.
DOI: 10.48550/arXiv.2601.11168.

End of Volume 6, Issue 3
Journal of Emerging Technologies and Computing (JETC)
Published by SDU University • © 2026
