

GEOSA

الهيئة العامة للمساحة
والمعلومات الجيومكانية

General Authority for Survey
and Geospatial Information



Designing AI Generative Agents for Arabic Geographical Names

Dr. Fatmah Baothman

AI Saudi Expert

some photos are generated by AI tools

Table of Contents



01

Opening: The main challenges affecting geographical Arabic names standardization and representation

02

The Challenge: Special Patterns Of Arabic Language Complexity and Cultural Sensitivity

03

The Solution: A Roadmap for Arabic Names Transcription Standardization Mechanism

04

How It Works: Guidelines & Case Study Standardization for Transcribing AI Geographical Arabic Names

05

AI Agents: Introduce & demonstrate different types of AIs to overcome challenges of Arabic names data processing

06

Multi-Agent System: Framework & Architectural Design for Multi-Agent Collaboration.

07

Use Case: Multi-Agent Systems for Arabic Geographical Names & performance of Sub symbolic representations.

08

Recommendations & Future Trends

Mapping the power of Arabic language & communication in a globalized world



1st Objective

To highlight the main challenges affecting Arabic geographical name standardization and representation.

Question to be addressed

How can artificial intelligence address the lack of clarity and support standardizing Arabic geographical names?



Solution Offered

Propose a sample of an IPA phonological transcription

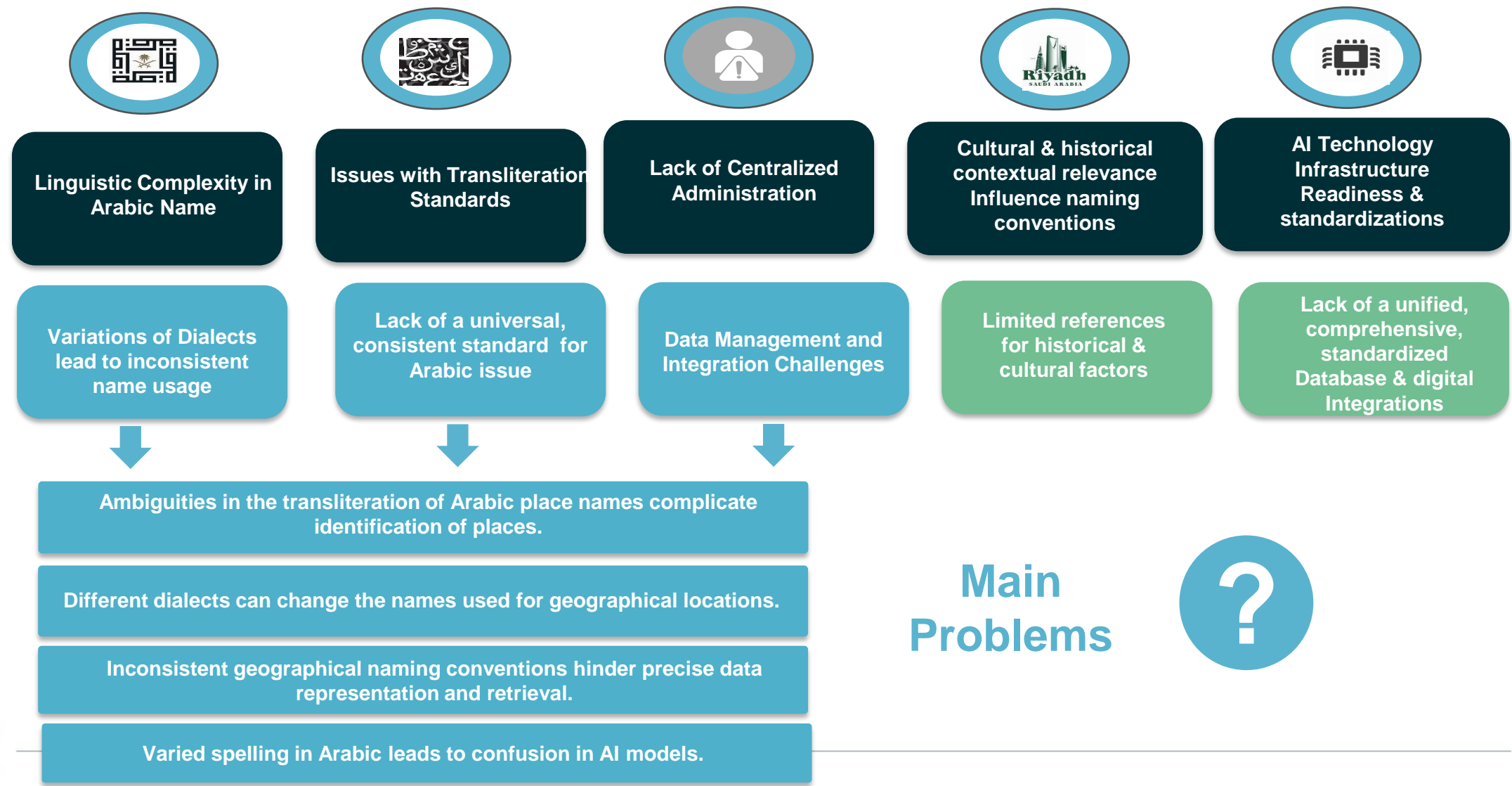


Benefits

Increased accuracy in data mapping and enhanced Arabic transcription



Challenges Facing Arabic Geographical Names : Complexity and Cultural Sensitivity



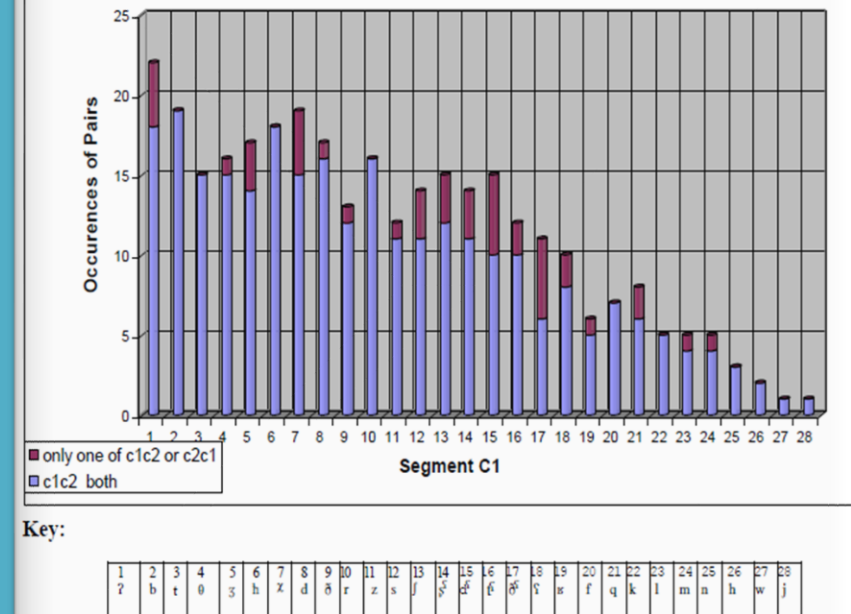
Special Patterns Of Arabic Language

1. All previous Arabic pattern studies focused on either classical Arabic or minority dialects
2. Arabic has more segmental contrasts than most languages (bɑrd ><bɑdr)
3. Representations are based on 6-syllable structures rather than one

Syllable Type	Conventional Syllabification	CV Syllabification	meaning	Arabic Words
1- CV	ka-ta-ba=CVCVCV	ka-ta-ba	he wrote	كَتَبَ
2- CVV	ʔaħ-za:-na-hu=CVCCVVCVCV	ʔa-ħø-za:-na-hu	his sadness	أَحْزَانُهُ
3- CVC	sʕum=CVC	sʕu-mø	fast!	صُم
4- CVVC	ri:f=CVVC	ri:-fø	cultivated land	رِف
5- CVCC	ħibr=CVCC	ħi-bø-rø	ink	حَبْر
6- CVVCC	χa:sʕ:=CVVCC	χa:-sʕø-sʕø	special	خَاصَّ

Traditional And CV Syllabification

Mirror Image Pairs and Non-Mirror Pairs in Arabic



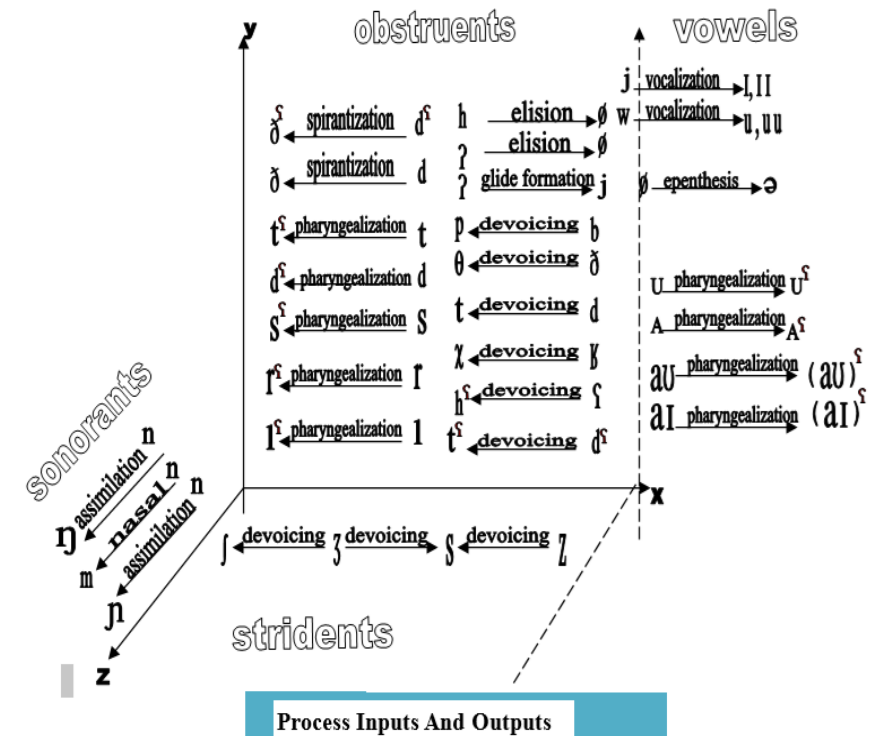
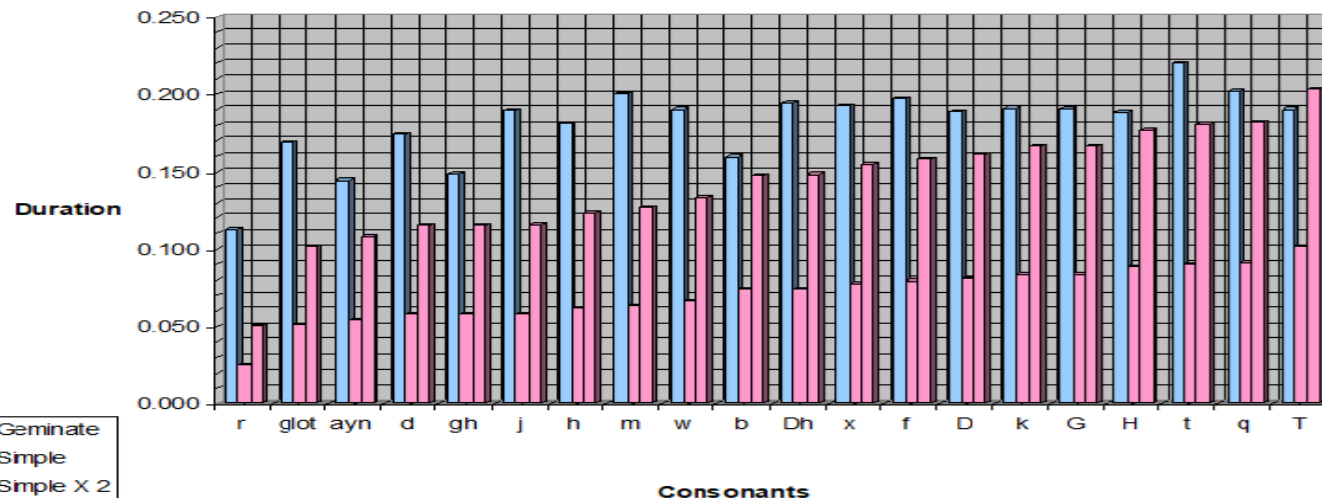
Distribution Of MIPs In Arabic Based On The AWC



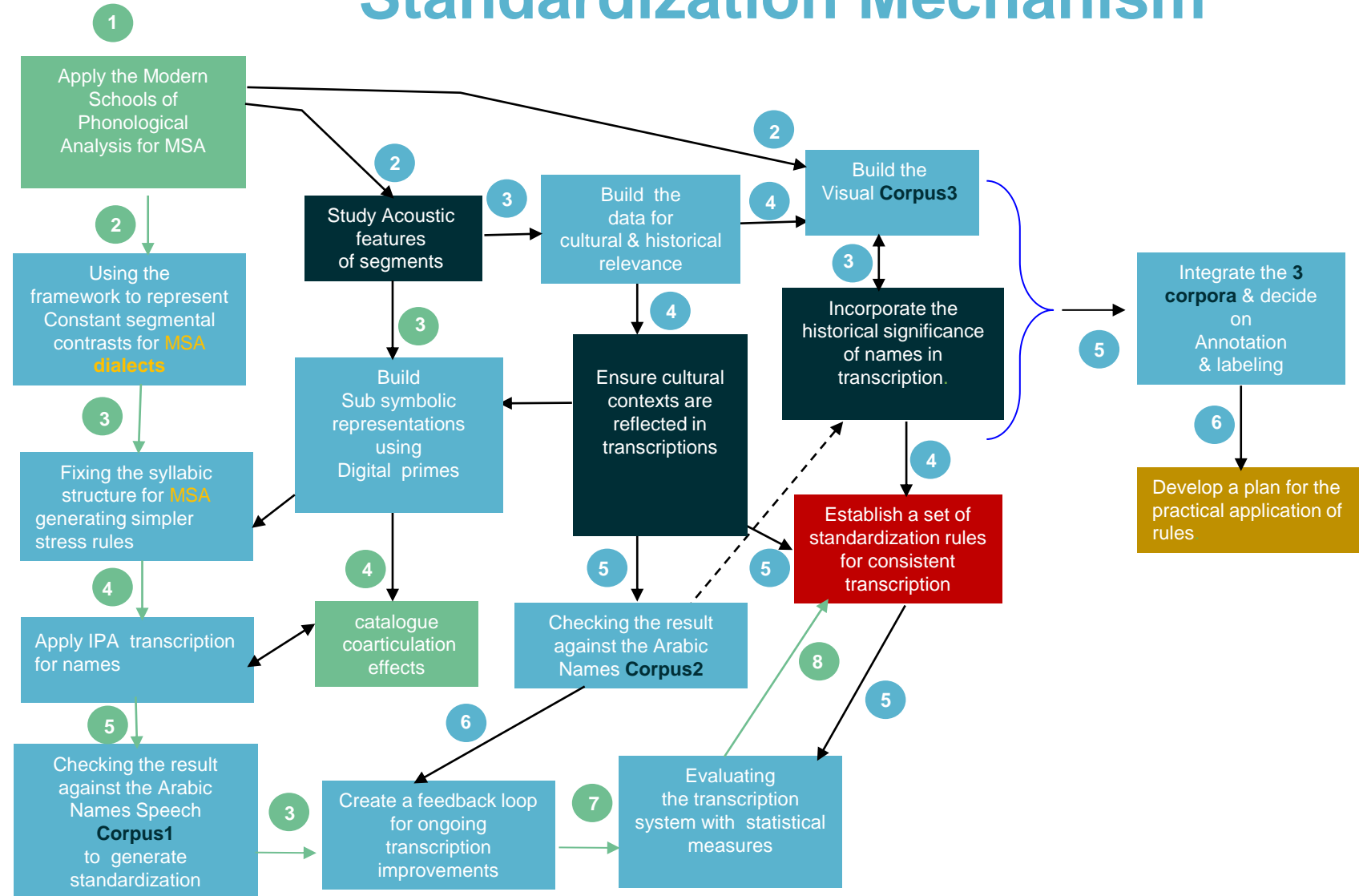
Special Patterns Of Arabic Language

4. Large number of Geminates
5. Phonological processes for MSA were not catalogued, nor mapped to their waveform representations and acoustic manifestations with IPA transcription
6. No unified phonological transcribed Corpus for MSA
7. Sub-symbolic representation had not been used to model Arabic Geographical Names

Average Length Geminate and Simple Consonants



A Roadmap for Arabic Names Transcription Standardization Mechanism



Guidelines for Transcribing Geographical Names



Spelling Consistency

Use a uniform transliteration scheme throughout the document.



Vowel & Geminales Representation

Use consistent symbols for Arabic vowel sounds in transcription.



Emphasis Marks

Indicate pharyngeal consonants with appropriate emphasis markers.



Dialect Variations

Account for regional dialects that affect sound production.



CV-Syllable Structure

Follow Arabic syllable structure to ensure natural flow in names.



Consonant Clustering

Represent clusters of consonants according to Arabic phonetics.



Sound Name Length

Transcribe long vowels distinctly to preserve their length.



Visual Representation

Employ clear visual indicators for different dialects.



Cultural Sensitivity Context

Respect cultural and regional variations in name pronunciation.



Historical Variants

Incorporate historical knowledge into naming modern transcription.



Consider IPA Phonetic Representation

Ensure that the phonetic elements are accurately represented in the transcription for clarity.



Evaluate and Measure the Standardized transcription accuracy

Established a set of guidelines for Arabic name transcriptions measures.



Case Study Standardization for AI Geographical Names



Problem Faced

Inconsistent transliterations of Arabic geospatial names across platforms.



Solution Offered

follow a guideline for an AI-driven system for transliteration standardization.



Benefits

Enhanced clarity, accuracy, and contextual relevance for cultural or historical data.

Arabic Romans	Arabic Names
Qalīb Ash <u>Shuhaybiyyah</u>	قَلِيب الشَّهْبِيَّة
Dahal <u>at</u> <u>Zimlah</u>	دَحَلَة زَمَلَة
Ar <u>Rūsān</u>	الرُّوسَان
<u>Khuff</u>	خُف
Al <u>Khufayfiyyah</u>	الْخُفَيْفِيَّة
Artāwī <u>Hillīt</u>	أرطاوي جَلِيت

Case Study for Transcription Sample

Name of letter	Isolated form of letter	IPA symbol (1993 version)	Suggested English transcription	Suggested HTK transcription	English sound	Phonetic description Maddieson 84	Arabic example	meaning	gemination form	meaning
30-Haa	هـ	h, h:	h	H	hair	glottal vd nonsib fricative	hawāʔ	air	muḥāʔah	developer (of a photo)
31-Waw	و	w, w:	w	W	wall	labial-velar vd cent approximant	wadi	valley	tasāwaw	conceit
32-Yaa	ي	i, j:	y	Y	yet	palatal vd cent approximant	jaktub	he writes	tājjar	airman
33-Hamza	ء	ʔ	ʔ	P	hour	glottal vl stop	ʔafḍalijah	advantage	saʔad	a quibbler
34-Alif-maqqura	ى	N/A	ay	AM	N/A	N/A	ramay	he threw	N/A	N/A
35-Taa-marbuʔah	ا	N/A	a(t)	TM	N/A	N/A	fakiha(t) al-musim	seasonal fruit	N/A	N/A

Table A Continued: The Phone Invento



2nd Objective

To demonstrate how semantic AI, actionable AI, agentic AI, and generative AI strengthen geographical names data processing.

Question to be addressed

What are the roles of semantic AI and generative AI in ensuring culturally informed and context-sensitive standardization?



Solution Offered

Developed AI generative multi-agent system to standardize geographical names and their representations.



Benefits

Increased accuracy in data mapping and enhanced evaluation results.





01

Semantic AI

Facilitating the representation of complex knowledge structures to enhance understanding of language meanings & accessibility for context.

Leveraging semantic relationships to enhance search engines and data retrieval in NLP

Used in building knowledge graphs for smarter data representation.

Enables the integration of diverse data sources

02

Actionable AI

Acts based on large data analytics, pattern understanding, and knowledge representations to provide immediate

categorize & classify tags into predefined classes for improved automated annotation & retrieval

Used for quick accurate translations across languages to automatically summarize large volumes of text for decision-making.

Enables real-time interactions by generating relevant answers based on text queries and patterns.



03

Agentic AI

Enables autonomous decision-making through adaptation, creating continuously self-improving systems.

Used for language personalization, huge datasets for enhanced machine understanding

Enables decision-making via text evaluation, independently without human interaction.

Leveraging NLP techniques allowing for enhanced problem-solving capabilities in complex tasks

04

Generative AI

Automate repetitive creative tasks to save time and resources to generate content dependent & engage in language writing, conversation, graphs..

Generates personalized text, multi-lingual translation, process complex tasks, improve contents and clarity,

Used for developing conversational agents, enhancing training datasets and improve outputs with iterations.

Enables summaries of documents, and Increases translation accuracy across multiple languages..

AI Concepts for Geographical Arabic names complexities

Framework for Multi-Agent Collaboration System

Multi-Agent Arabic Names System Overview

Agent Design

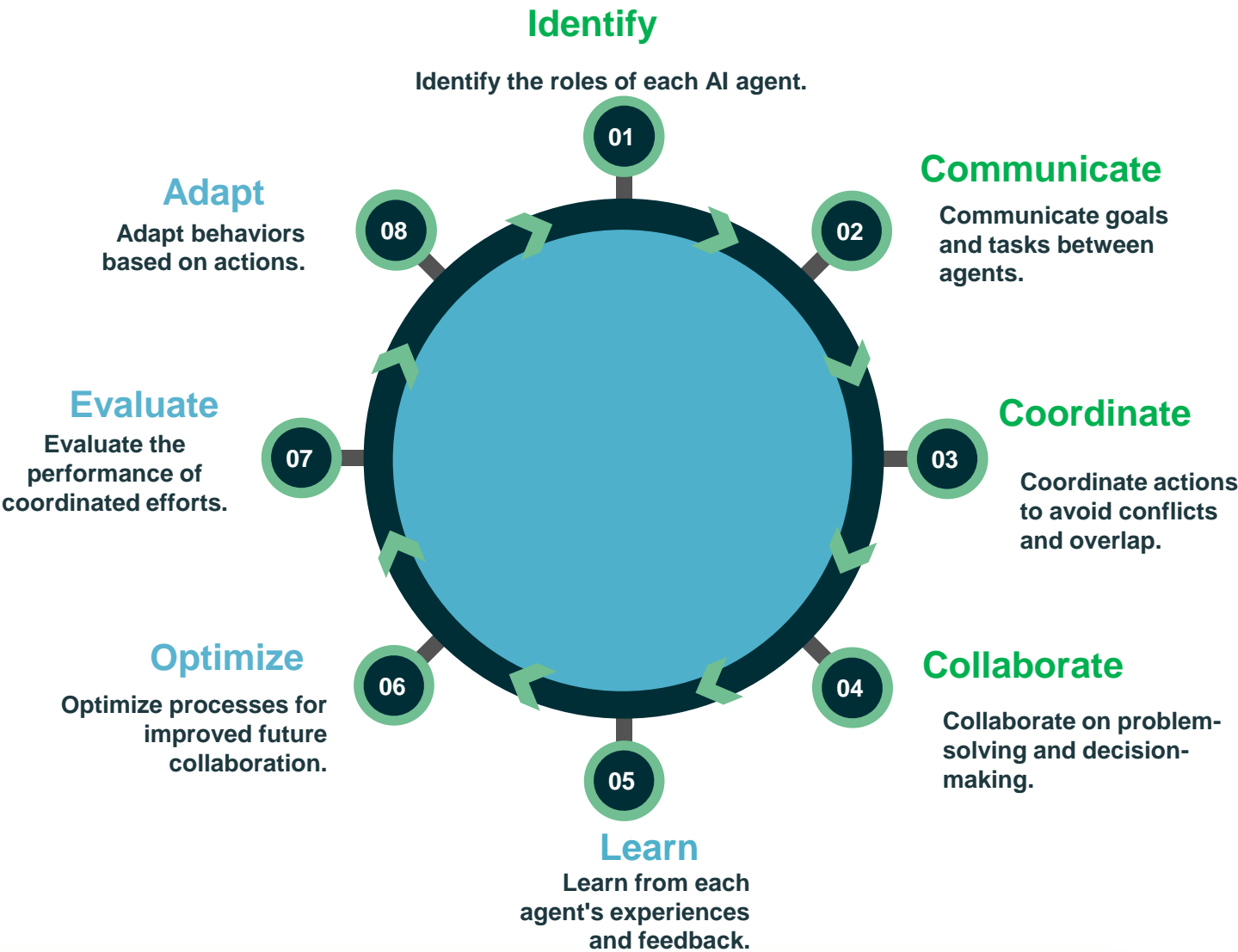
Focus on Arabic geographical name recognition and generation.

Collaboration Strategy

Define interaction patterns among agents for efficiency.

Data Management

Handle and store geographical names data effectively.



Description

Environments

Perception

Actuator

Contextualization

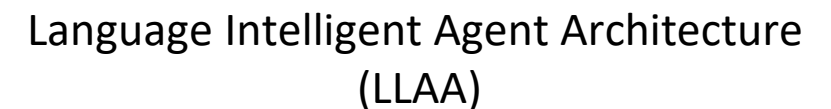
Agent Coordination

Data Validation

Semantic Unit

Output

**Generate
standardized
geographical name
outputs efficiently.**



3rd Objective

To present an AI-driven, multi-agent system solution for addressing linguistic transliteration for different Arabic languages countries & dialects

Question to be addressed

What challenges remain in applying AI models across different Arabic dialects, and how can they be mitigated?



Solution Offered

Present a Use Case for an AI multi-agent system dealing with Arabic geographical names and dialects.



Benefits

Data mapping and enhanced results for geographical representations of Arabic names



Multi-Agent Architecture Design for Arabic Geospatial Names

Agent Skills

Form Chat LLM API GPT Postman

Select one of your agent's skills from the list, then click Test.

Transliterate

Transliterate

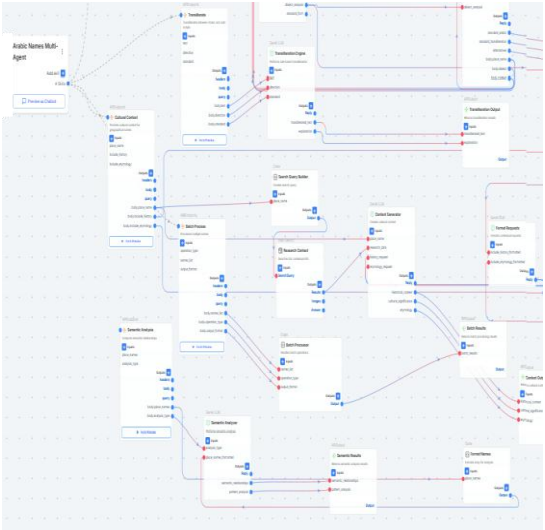
Cultural Context

Batch Process

Semantic Analysis

Standardize Name

Test



Test as Chat

Form Chat LLM API GPT Postman

Chat with your agent using an interactive chatbot UI. This can be used for live testing or embedded into your website.

Hi, I am Arabic Names Multi-Agent. How can I help you?

طريق وشارع
write all the Arabic Names for

The Arabic terms for "طريق" (road) and "شارع" (street) can have several variants and usage contexts. Here are several names and synonyms for each in Arabic:

For طريق (Road):

1. طريق - Road
2. درب - Path
3. مسار - Track/Route
4. خط - Line (in the context of a transport line)
5. ناحية - Area/District (in some contexts refers to a road)
6. ممر - Passage

For شارع (Street):

1. شارع - Street
2. زقاق - Alley
3. طريق - Can sometimes refer to a street in certain contexts
4. ميدان - Square (while not exactly a street, it often encompasses street areas)
5. حارة - Lane (sometimes implies a street or neighborhood passage)
6. ممر - Walkway/Passage (lesser-used synonym)

These terms can vary in different Arabic-speaking countries, and local dialects may have specific names or usages that differ from Modern

The Arabic name "الروسان" can be transliterated into Latin script in various ways depending on the standard used. Here are the variations based on two common transliteration standards:

1.ALA-LC (American Library Association - Library of Congress) DIN rof etutitsnl namreG)31635 (noitazidradnatS

2.Summary of Variations is

- al-Rūsān
- ar-Rawsān

These forms preserve the phonetic qualities of the original Arabic name and can be used depending on the context or audience requiring a specific standard .

Approach

01

Define Agent

Gather diverse Arabic geographical names data.

02

AI Integration

Utilize semantic AI for better contextual understanding.

03

Multi-Agent Design

Create agents to manage different naming tasks.

04

Validation Process

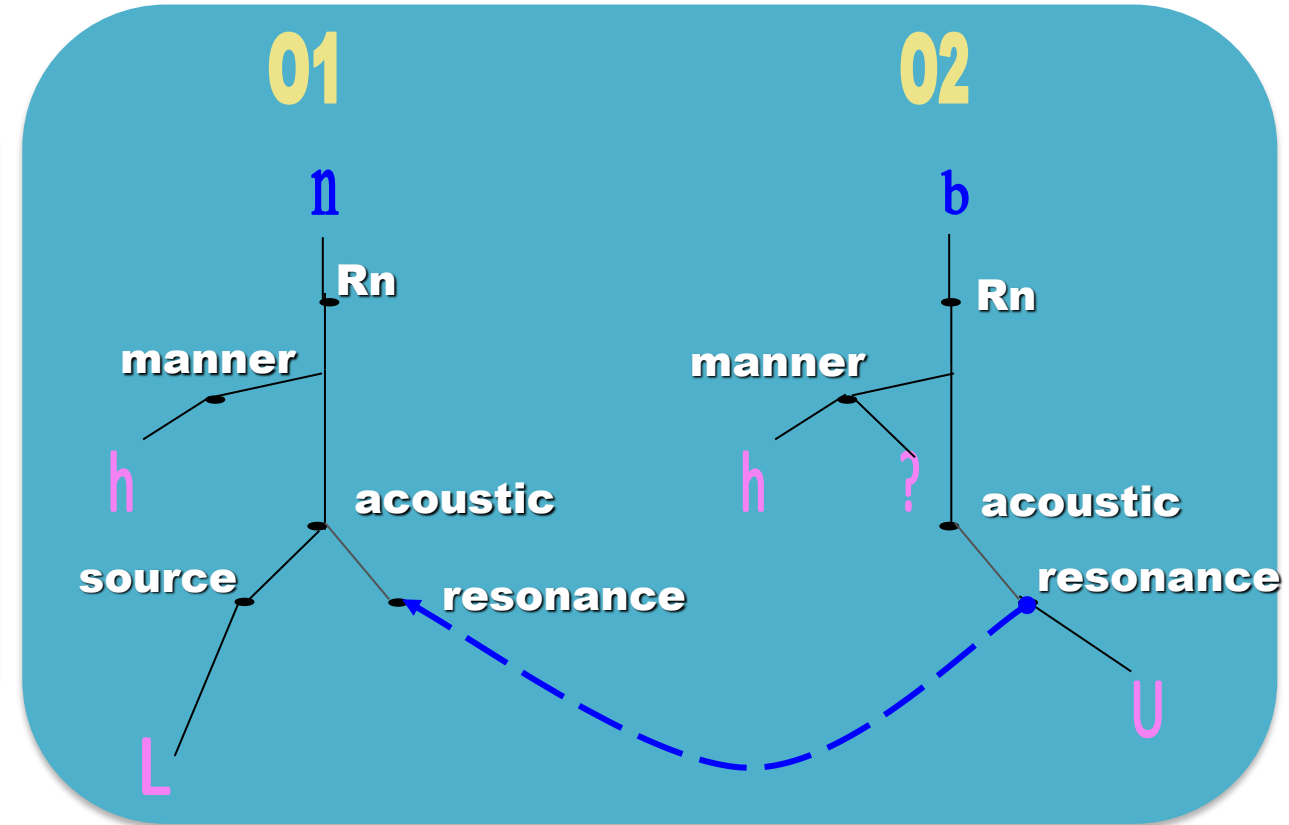
Implement checks to ensure name accuracy & variations.



Sub-symbolic modelling coarticulation effects in Arabic

manner		p	b			s	s ⁹			m	n			w	j
	?	1	1												
resonance	h	1	1			1	1			1	1			1	1
	U	1	1							1				1	
	I					1	1								1
source	A						1								
	H	1				1	1								
	L						1			1	1				

Agent Representation of short range minber → mimber



Symbolic VS. Sub-symbolic Performance

Aspect	Symbolic AI	Sub symbolic AI
Representation	Rule-based system	Learned patterns from data
Value	Effective for structured tasks	Handles regional dialects

Training

Subsymbolic

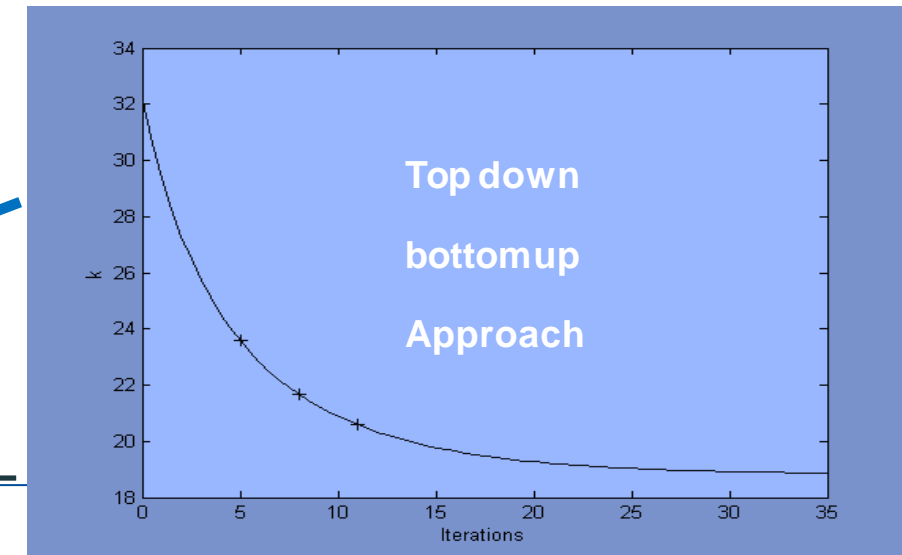
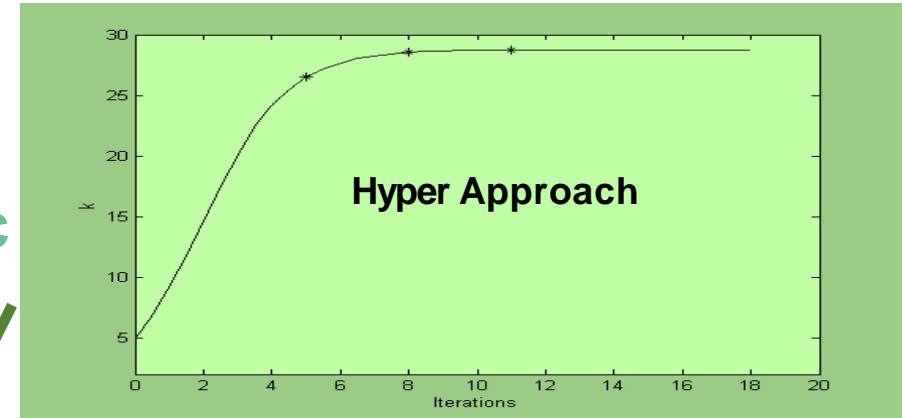
Symbolic

1940 1950 1960 1970 1980 1990 2000 2010



Timeline

Henry Lieberman - MIT



Fatmah Baothman – KSA (2003-2025)

Recommendations & Future Trends



01

Adapt AI next generation transcription for Arabic geographical names

02

leverage generative AI for standardizing geographical names across Arabic-speaking region preserving cultural & heritage

03

Align AI with the UNGEGN (United Nations Group of Experts on Geographical Names) framework

The Mapping Symbolizes the interconnected Arabic language, culture and heritage globally toward AI Design Principles



Thank You



شكراً لكم



Fatmah@ai4ia.tech
fbaothman@kau.edu.sa

info@geosa.gov.sa
geosa.gov.sa

GEOSA
الهيئة العامة للمساحة
والمعلومات الجيومكانية
General Authority for Survey
and Geospatial Information

