

**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



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



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

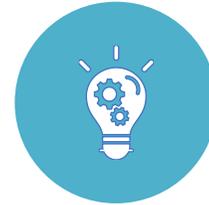
**Recommendations & Future Trends**

## 1<sup>st</sup> 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

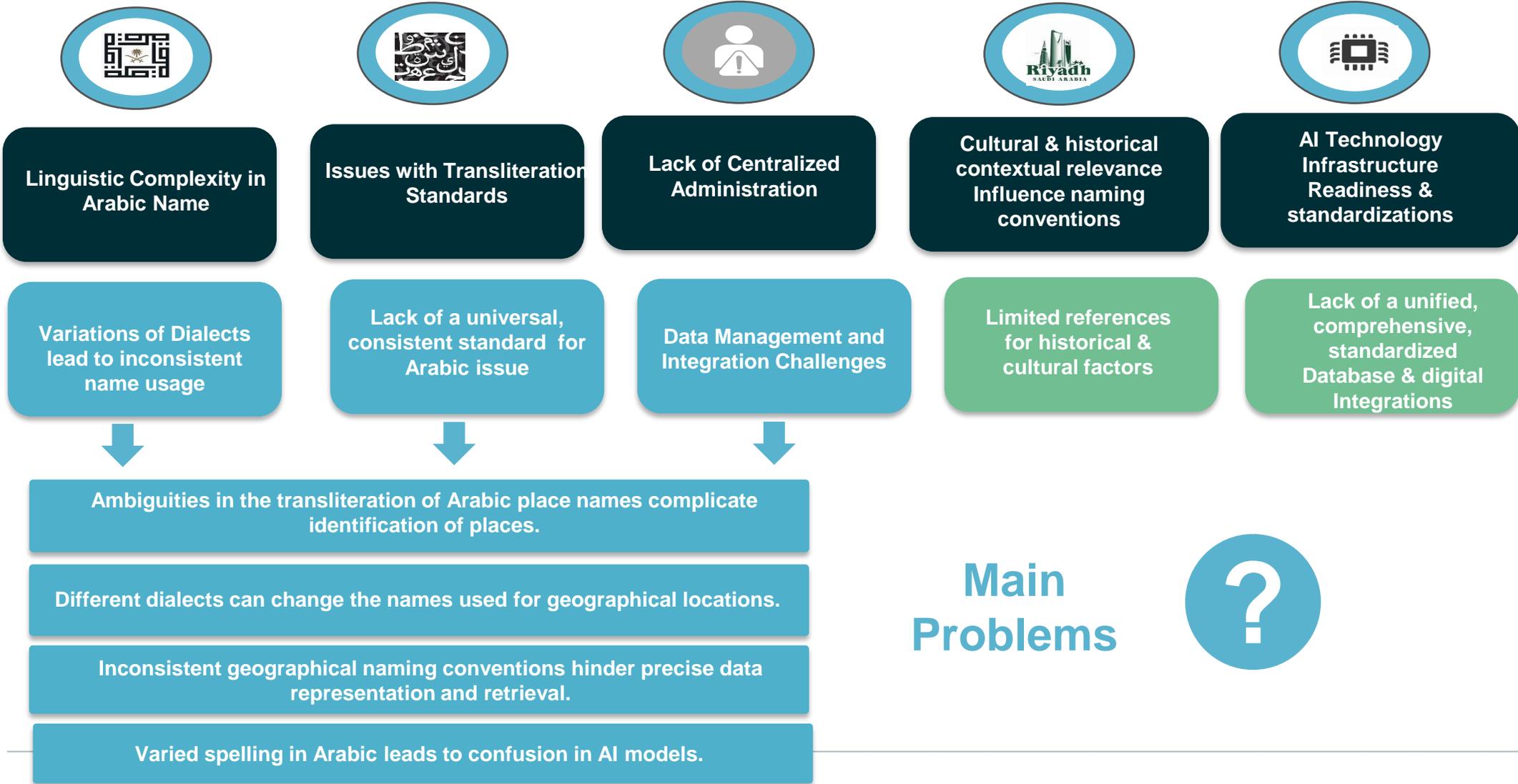


## Benefits

Increased accuracy in data mapping and enhanced Arabic transcription



# Challenges Facing Arabic Geographical Names : Complexity and Cultural Sensitivity



# Special Patterns Of Arabic Language

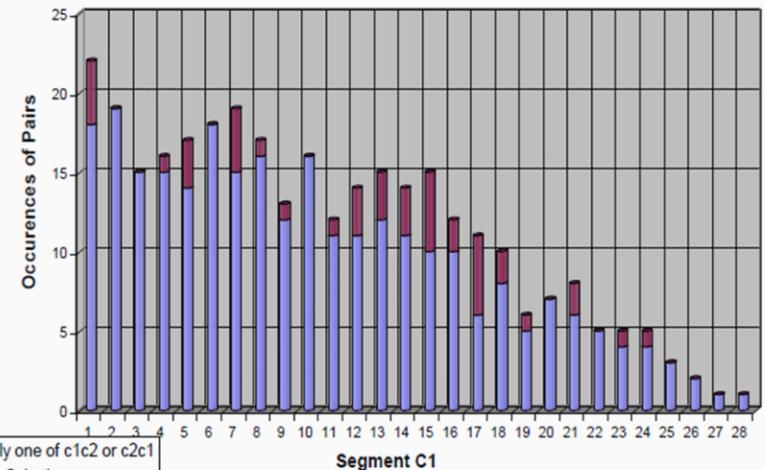
1. All previous Arabic pattern studies focused on either classical Arabic or minority dialects
2. Arabic has more segmental contrasts than most languages (bard ><badr)
3. Representations are based on 6-syllable structures rather than one

Syllable Type	Conventional Syllabification	CV Syllabification	meaning	Arabic Words
1- CV	ka-ta-ba=CVCVCV	ka-ta-ba	he wrote	كَتَبَ
2- CVV	?aḥ-za:-na-hu=CVCCVVCVCV	?a-ḥø-za:-na-hu	his sadness	أَحْزَانُهُ
3- CVC	s <sup>ʕ</sup> um=CVC	s <sup>ʕ</sup> u-mø	fast!	صُم
4- CVVC	ri:f=CVVC	ri:-fø	cultivated land	رِيف
5- CVCC	ḥibr=CVCC	ḥi-bø-rø	ink	حَبْر
6- CVVCC	χa:s <sup>ʕ</sup> =CVVCC	χa:-s <sup>ʕ</sup> ø-s <sup>ʕ</sup> ø	special	خَاصَّ

## Traditional And CV Syllabification



Mirror Image Pairs and Non-Mirror Pairs in Arabic



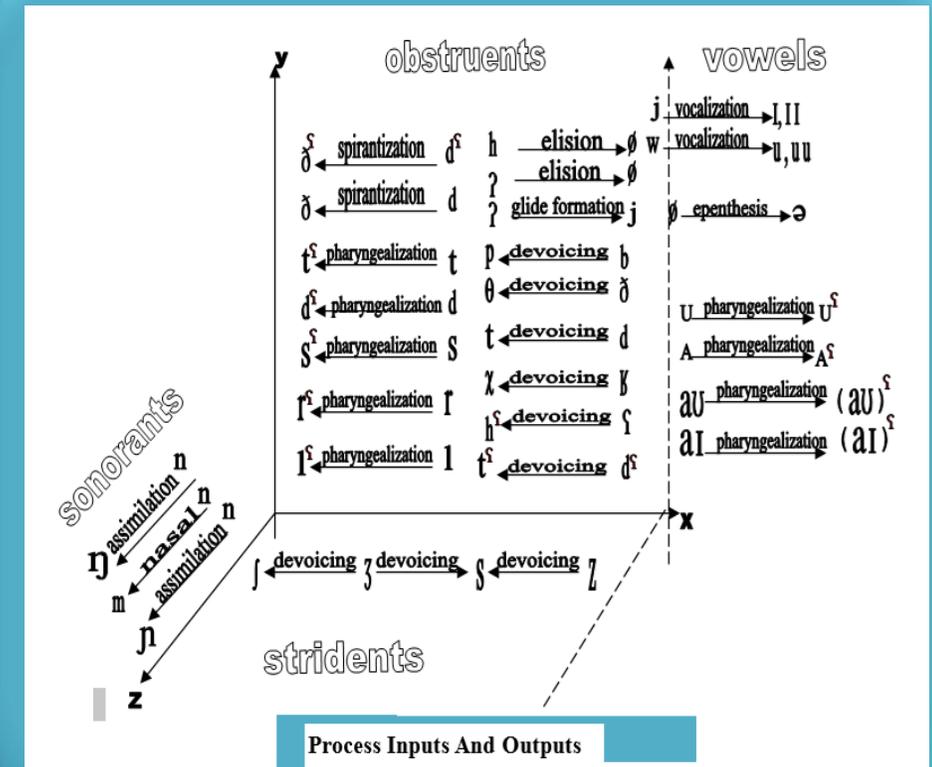
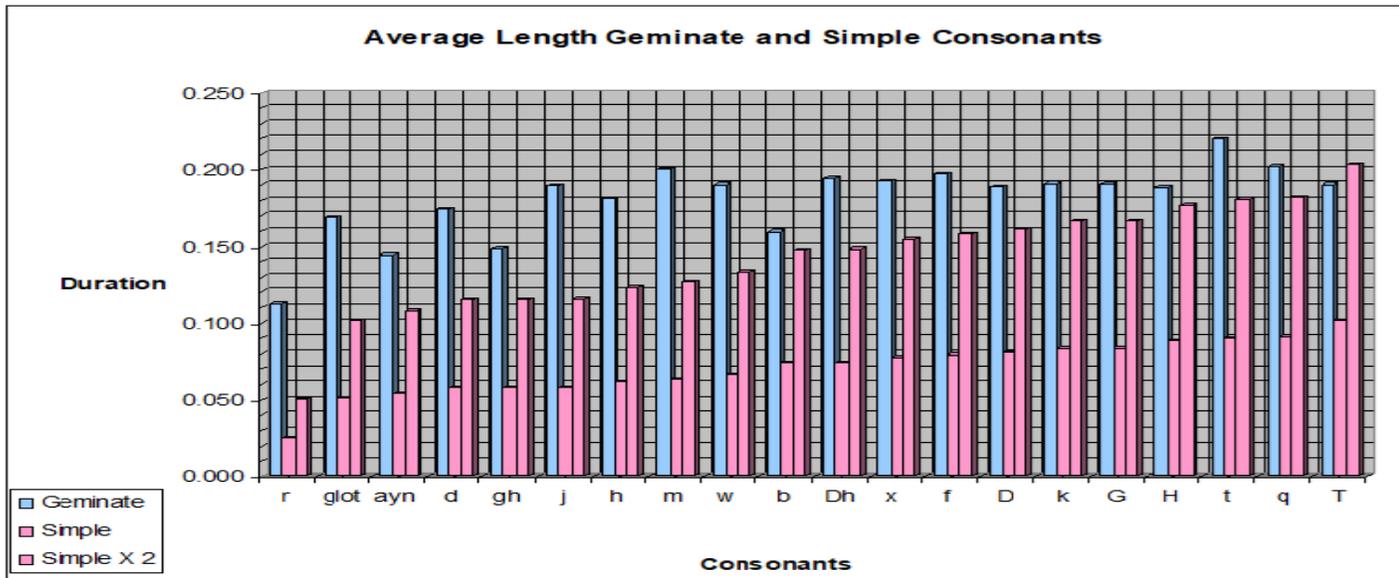
Key:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
?	b	t	θ	s	h	x	d	r	z	s	j	s <sup>ʕ</sup>	d <sup>ʕ</sup>	f <sup>ʕ</sup>	ð <sup>ʕ</sup>	ʕ	s	f	q	k	l	m	n	h	w	j	

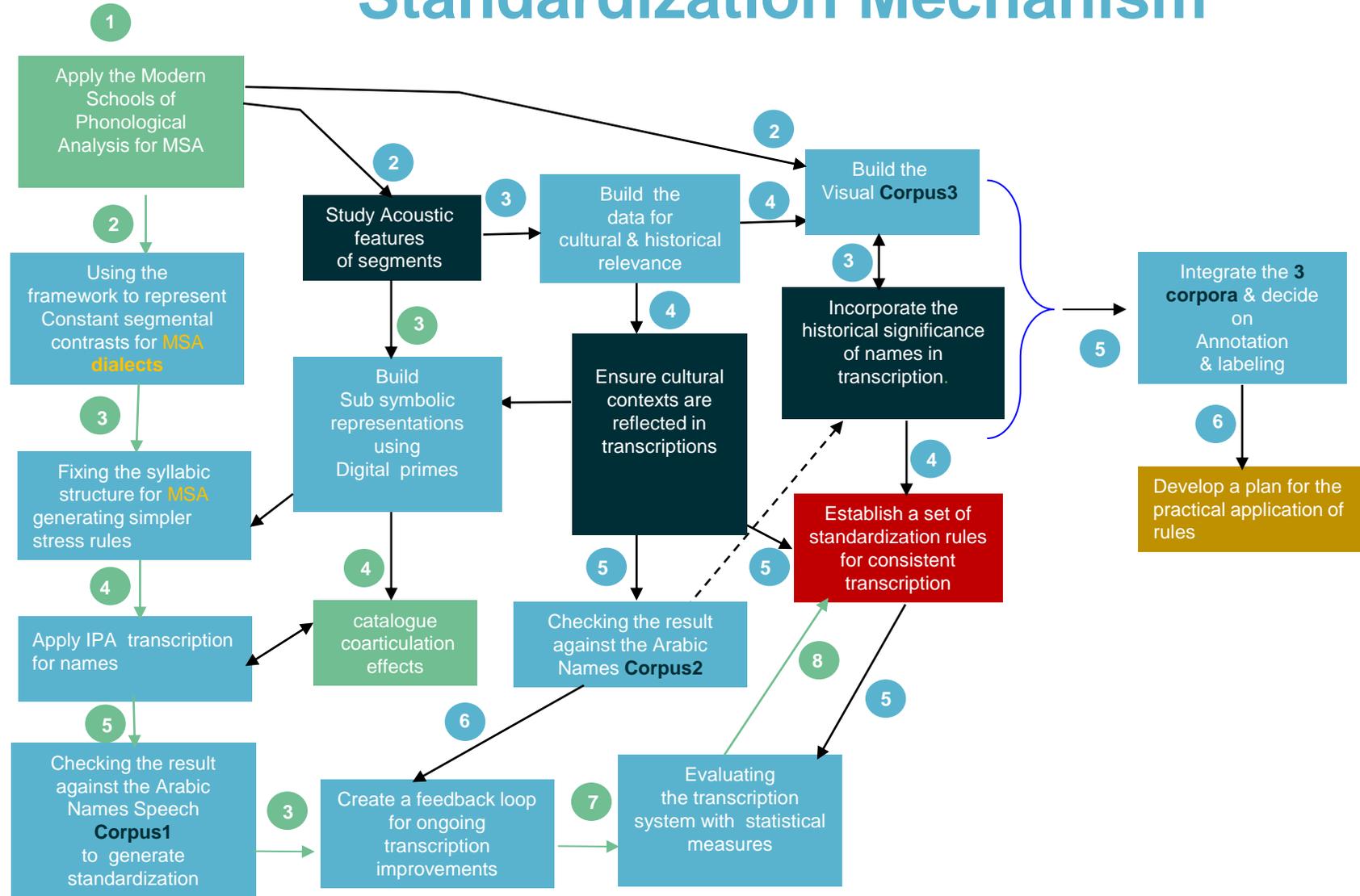
Distribution Of MIPs In Arabic Based On The AWC

# 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



# A Roadmap for Arabic Names Transcription Standardization Mechanism



# Guidelines for Transcribing Geographical Names



## Spelling Consistency

Use a uniform transliteration scheme throughout the document.



## Vowel & Gemimates 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
<u>Qalīb Ash Shuhaybiyyah</u>	قَلِيب الشَّهْبِيَّة
<u>Dahalat Zimlah</u>	دَحَلَة زِمْلَة
<u>Ar Rūsān</u>	الرُّوسَان
<u>Khuff</u>	خُف
<u>Al Khufayfiyyah</u>	الْخُفَيْفِيَّة
<u>Artāwī 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	ا	h, h:	h	H	hair	glottal vd.nonsib fricative	hawāʔ	air	muḥʔahir	developer (of a photo)
31-Waw	و	w, w:	w	W	wall	labial-velar vd.cent approximant	wadi	valley	tasʔawwir	conceit
32-Yaa	ي	j, j:	y	Y	yet	palatal vd.cent approximant	jaktub	he writes	tʔajjar	airman
33-Hamza	ء	ʔ	ʔ	P	hour	glottal vl stop	Zafdʔalijah	advantage	saʔrad	a quibbler
34-Alif-maqsura	ى	N/A	ay	AM	N/A	N/A	ramay	he threw	N/A	N/A
35-Taa-marbuʔah	آ	N/A	a(t)	TM	N/A	N/A	fakiha(ʔ) al-musim	seasonal fruit	N/A	N/A

Table A Continued: The Phone Invento



## 2<sup>nd</sup> 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

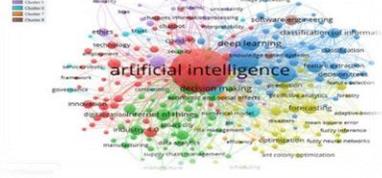
Increased accuracy in data mapping and enhanced evaluation results.



# AI Types & Agents Simplified

## Semantic AI

Focuses on understanding and interpreting human language in context.



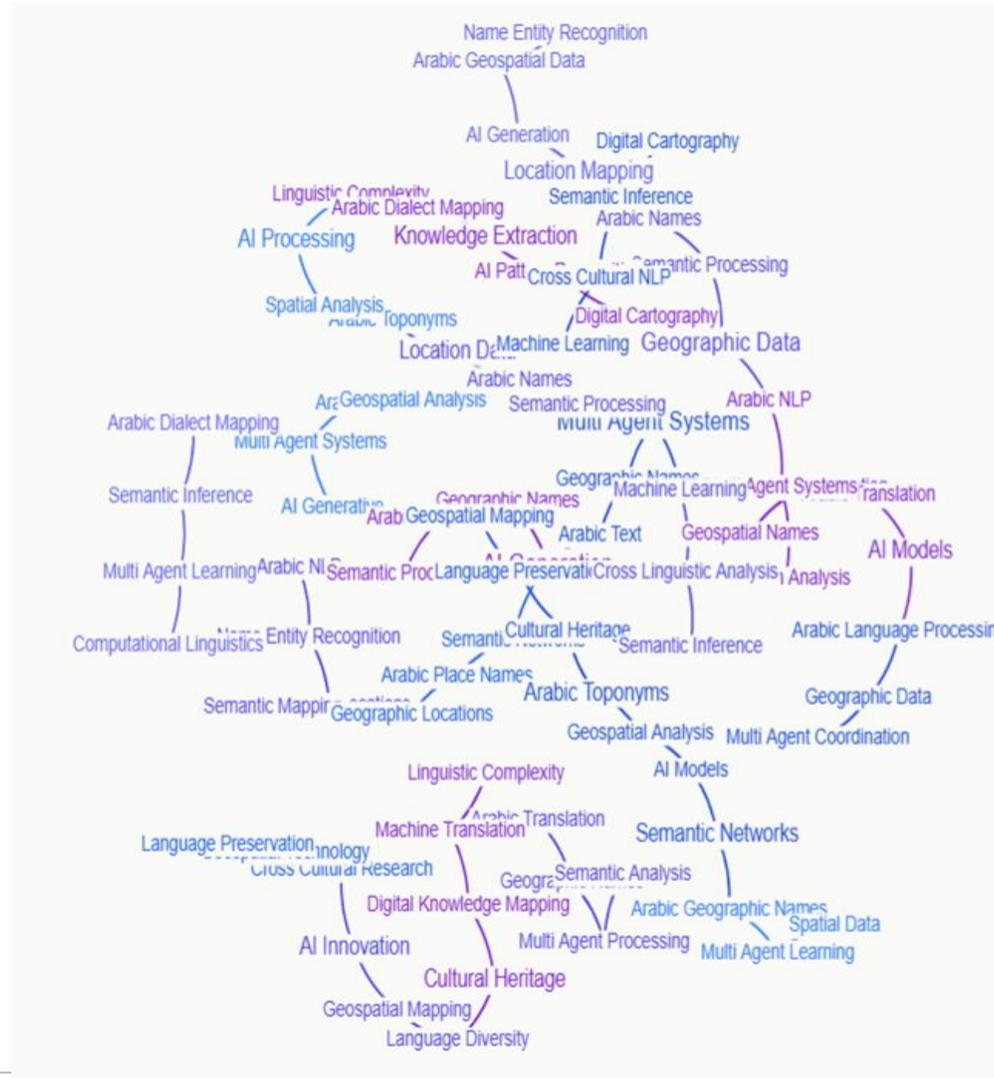
## Agentic AI

Represents AI that can operate independently in dynamic environments.



## AI Agent

Autonomous entities that can perform tasks on behalf of users for a specific task in a defined environment.



## Actionable AI

Enables decision-making with real-time data analysis and insights.



## Generative AI

Creates new content or solutions based on learned patterns and data.



## AI Multi-Agents

Systems where multiple AI agents collaborate to solve complex problems to achieve



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.



# AI Concepts for Geographical Arabic names complexities

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..

# Framework for Multi-Agent Collaboration System

## Multi-Agent Arabic Names System Overview

### Agent Design

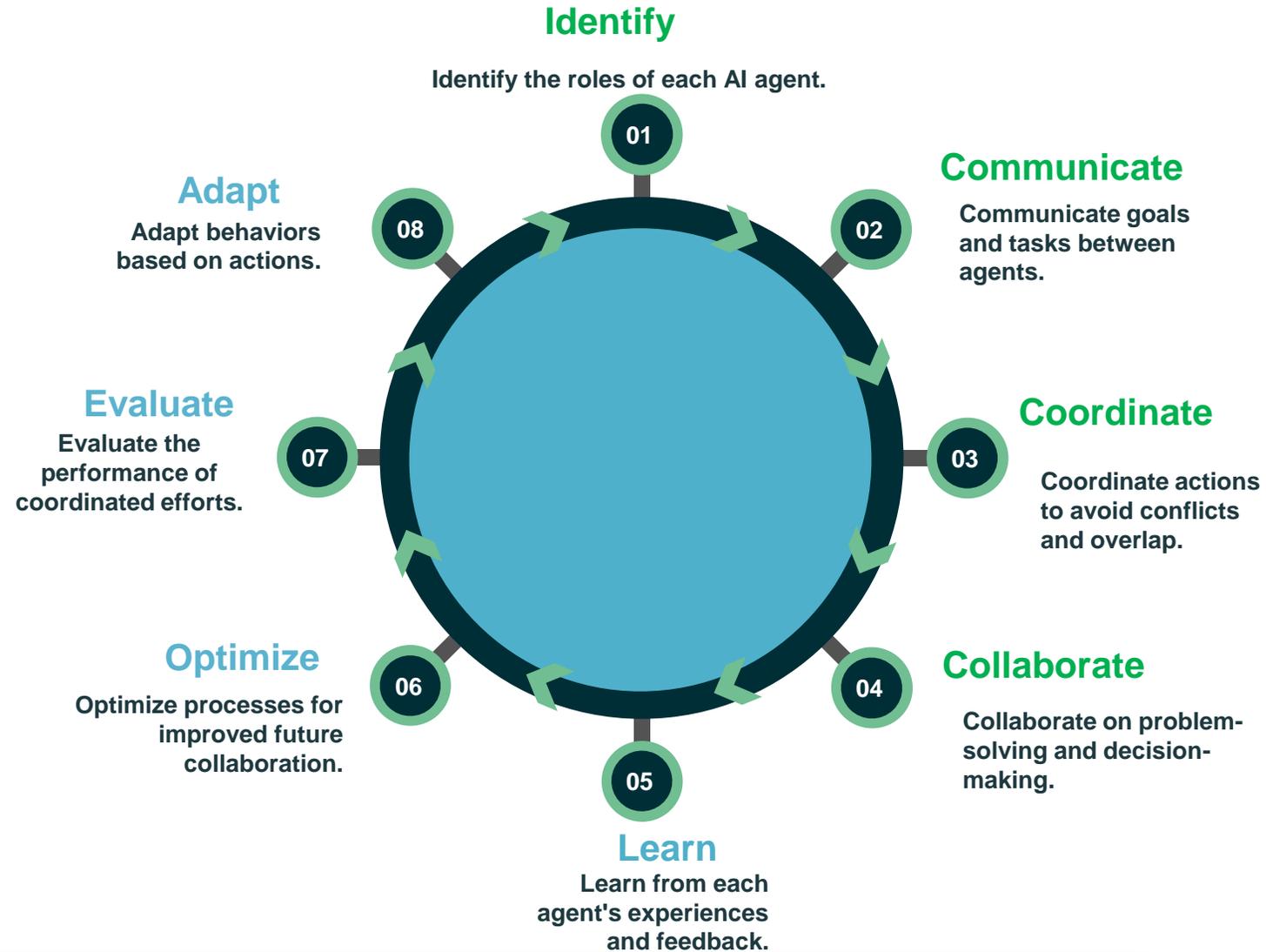
Focus on Arabic geographical name recognition and generation.

### Collaboration Strategy

Define interaction patterns among agents for efficiency.

### Data Management

Handle and store geographical names data effectively.



# Multi-Agent Architecture Design for Arabic Geospatial Names

## InformationwOLF

This structure represents a comprehensive agent architecture that combines perception, reasoning, decision-making, communication and environments capabilities in a unified framework.

### Description

01

#### Environments

collect diverse Arabic geographical names to create datasets.

02

#### Perception

Information is perceived, represented as knowledge, and reasoning.

03

#### Actuator

Decisions are made to convert names into standardized phonetic / digital representations.

04

#### Contextualization

Embed cultural/historical and geographical context.

05

#### Agent Coordination

Enable agents to collaborate and communicate with other agents in the environment for optimal results.

06

#### Data Validation

Language components support both understanding and expressions.

07

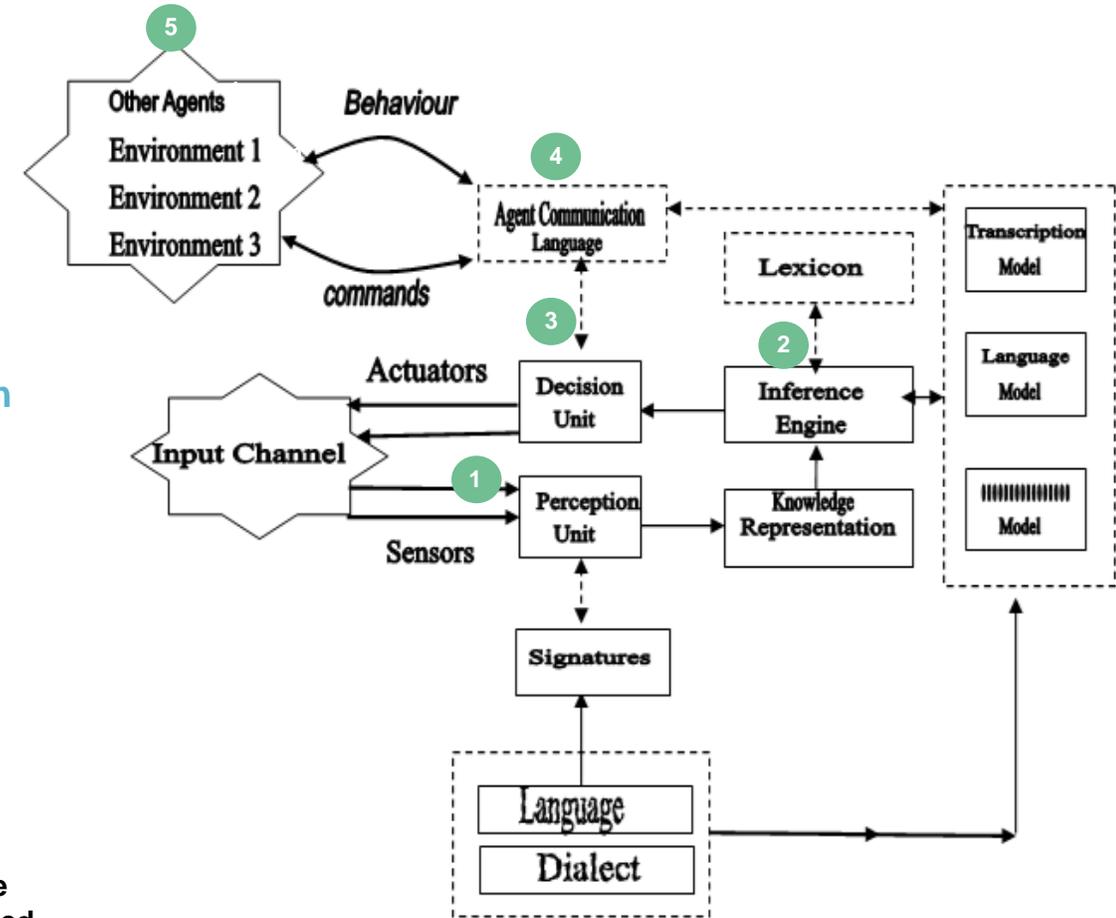
#### Semantic Unit

Enhance data through semantic understanding methods.

08

#### Output

Generate standardized geographical name outputs efficiently.



Language Intelligent Agent Architecture (LLAA)



## 3<sup>rd</sup> 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



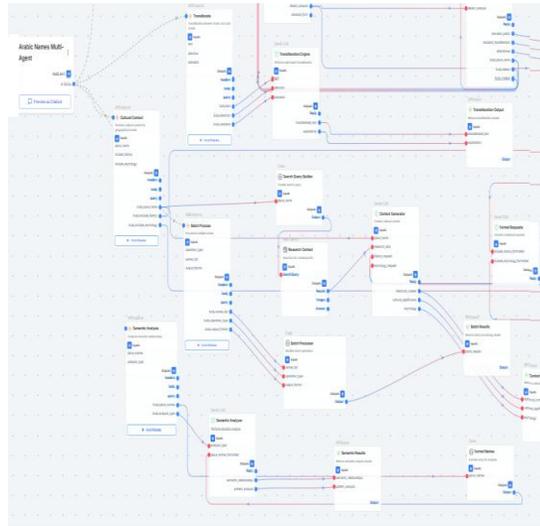
# Multi-Agent Architecture Design for Arabic Geospatial Names

## Agent Skills

Form Chat LLM API GPT Postman

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

- Transliterate
- Cultural Context
- Batch Process
- Semantic Analysis
- Standardize Name



## 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?

write all the Arabic Names for طريق وشارع

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):**

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

**For شارع (Street):**

- شارع - Street
- زقاق - Alley
- طريق - Can sometimes refer to a street in certain contexts
- ميدان - Square (while not exactly a street, it often encompasses street areas)
- حارة - Lane (sometimes implies a street or neighborhood passage)
- 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

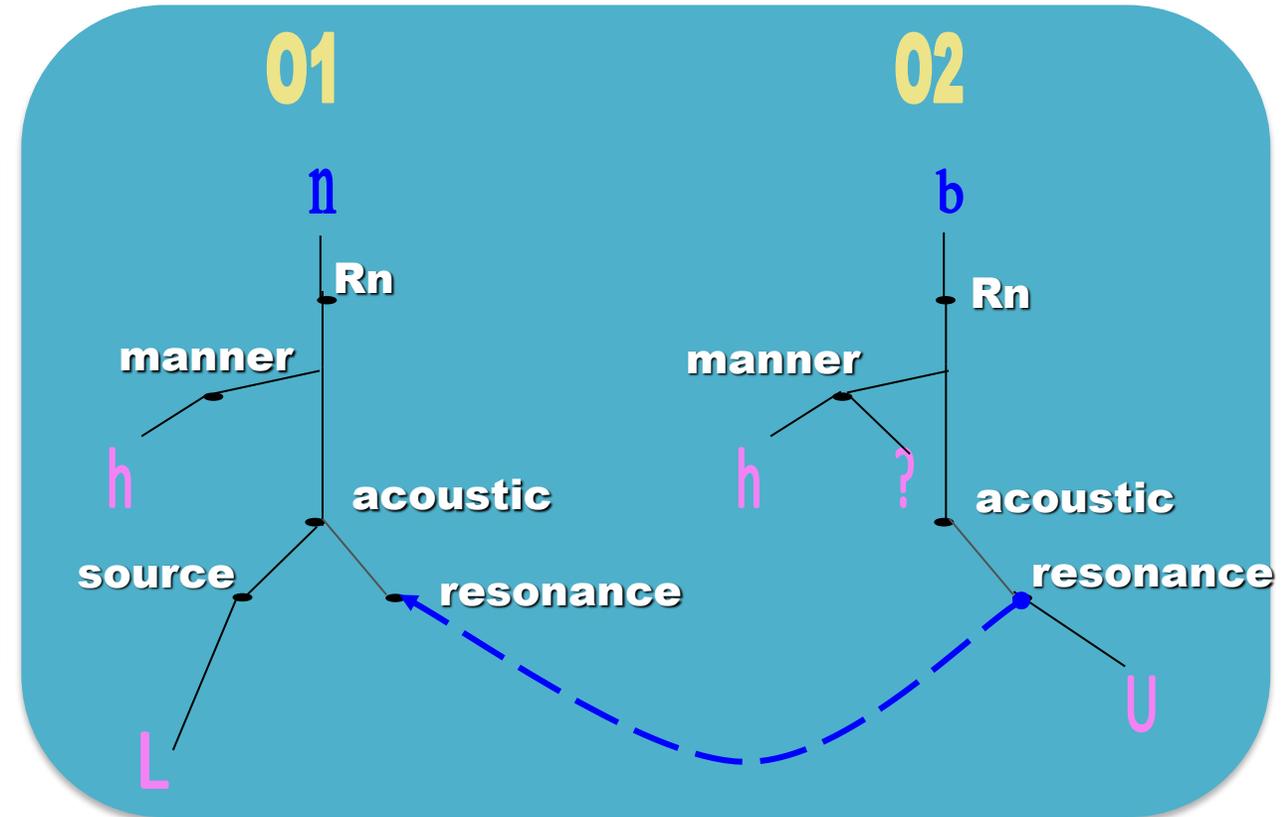
Implement checks to ensure name accuracy & variations.



# Sub-symbolic modelling coarticulation effects in Arabic

manner		p	b	s s <sup>9</sup>		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

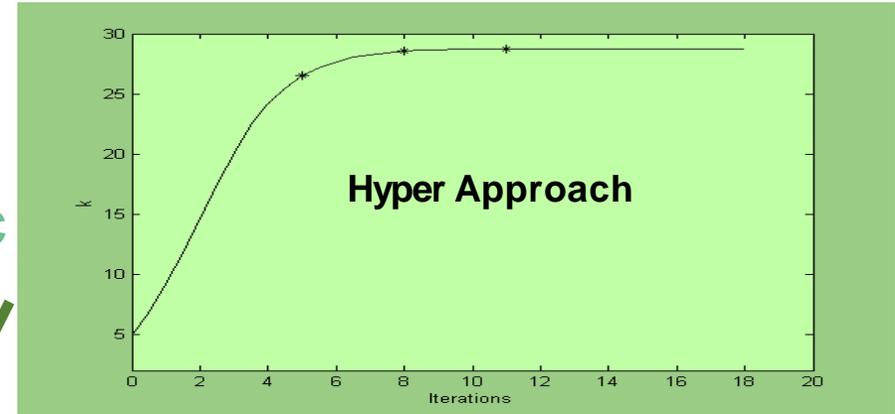


# Symbolic VS. Sub-symbolic Performance

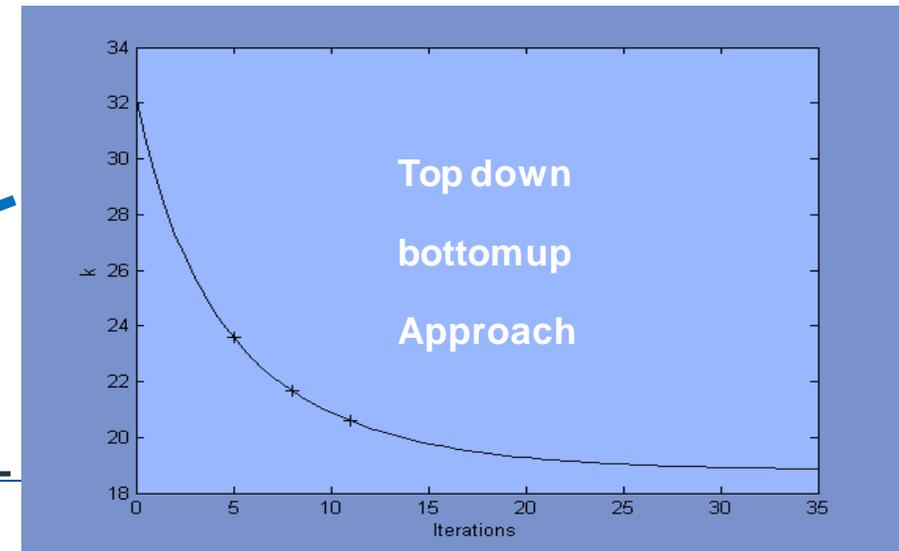
Aspect	Symbolic AI	Sub symbolic AI
Representation	Rule-based system	Learned patterns from data
Value	Effective for structured tasks	Handles regional dialects

Training

Subsymbolic



Symbolic



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](mailto:info@geosa.gov.sa)  
[geosa.gov.sa](http://geosa.gov.sa)

**GEOSA**

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

