

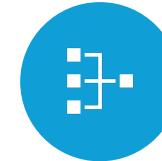
# Contents



Motivations



Research



System



Expert  
System



Adaptiveness



Other  
Achievements



Conclusions

Text ‘S’ for Smallholding

# Text ‘S’ for Smallholding

A Personalised Unified SMS-based Agricultural Extension Service  
for Small-Scale Farmers

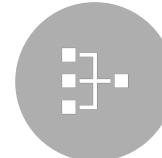
# Contents



Motivations



Research



System



Expert  
System



Adaptiveness



Other  
Achievements



Conclusions

# Personal Motivations



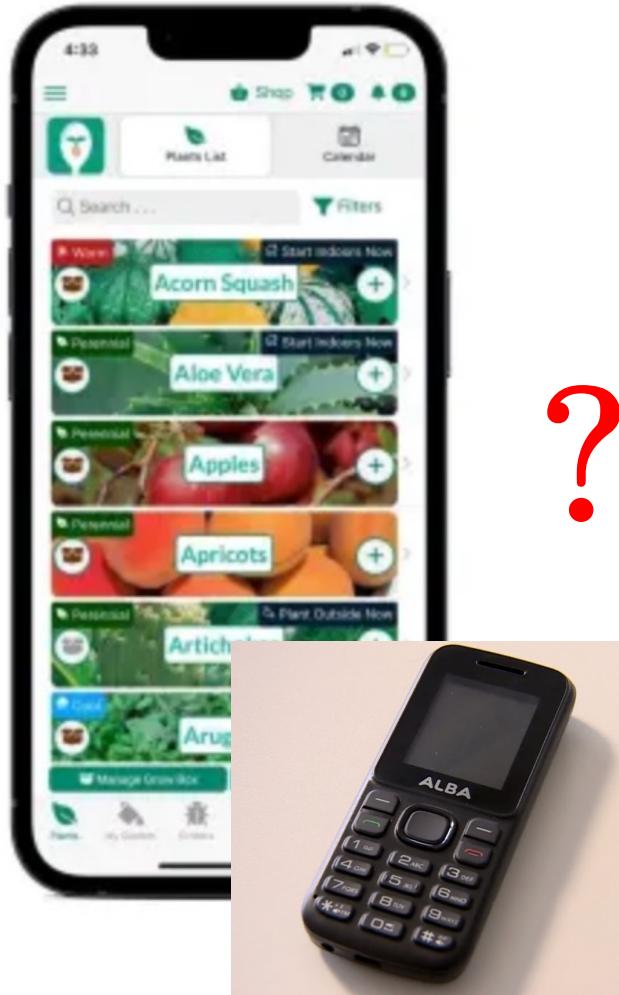
# Personal Motivations



# Personal Motivations



# Personal Motivations

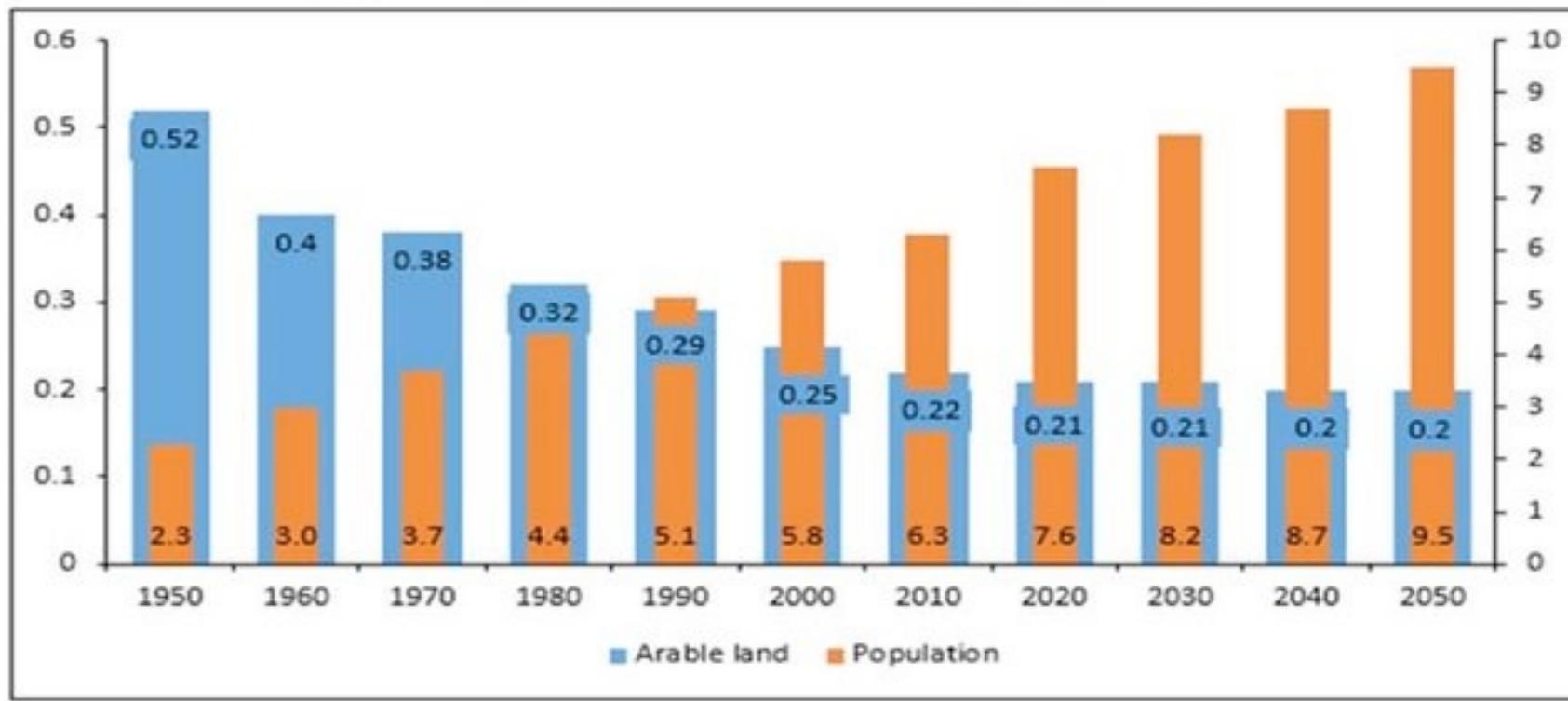


?



# Motivations

World population (billions) versus arable land (ha per person) 1950-2050



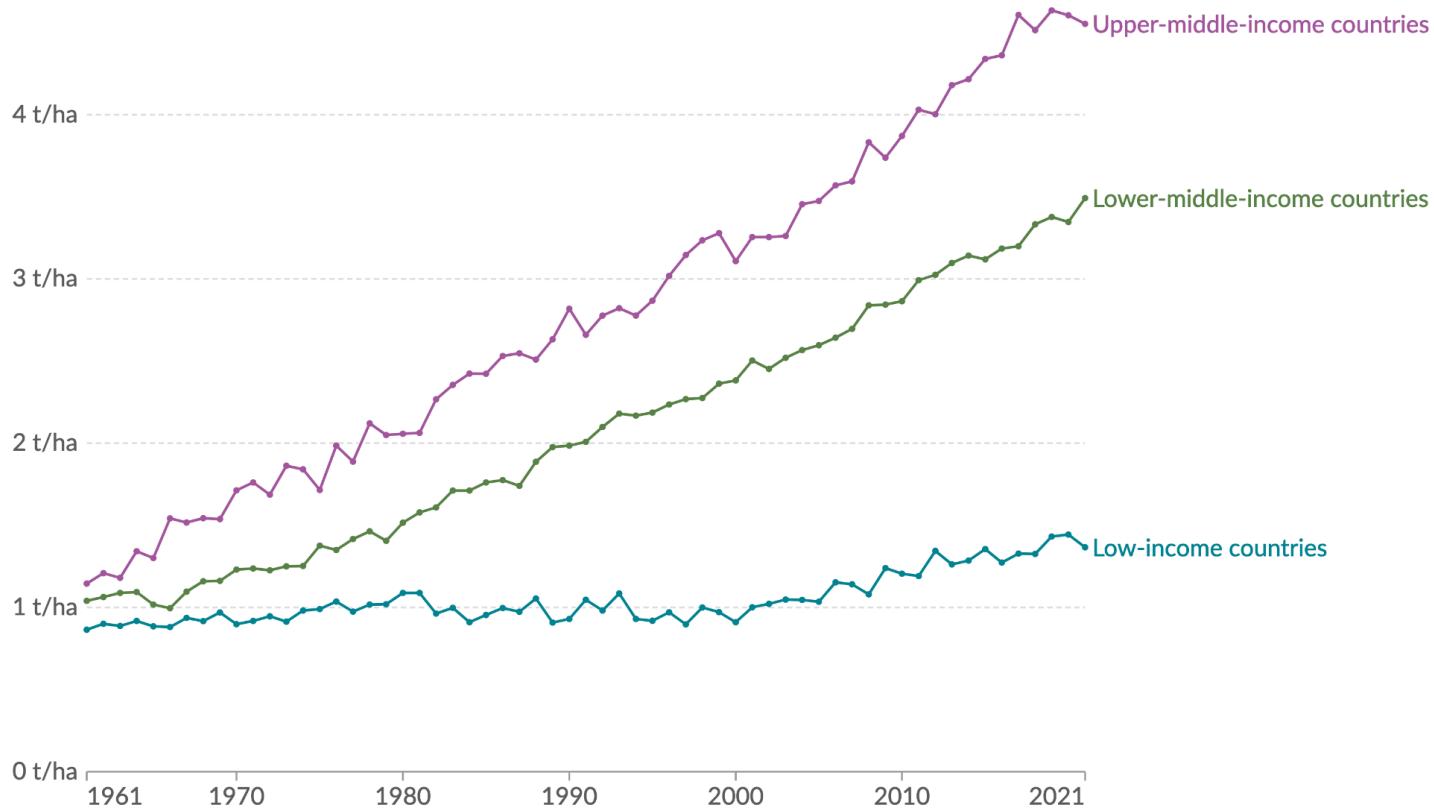
Source: Technavio, 2015

# Motivations

## Cereal yield, 1961 to 2021

Yield is measured as the quantity produced per unit area of land used to grow it.

Our World  
in Data



Data source: UN Food and Agriculture Organization (FAO)

Note: Cereals include wheat, rice, maize, barley, oats, rye, millet, sorghum, buckwheat, and mixed grains.

CC BY

# Terminology

What is  
Agriculture 4.0?

- IoT with Agriculture
- Using real-time data for insights
- ‘Precision Agriculture’

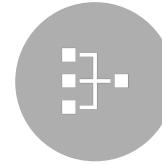
# Contents



Motivations



Research



System



Expert  
System



Adaptiveness



Other  
Achievements



Conclusions

# Terminology

What is an AES?

- Agricultural Extension Service
- Educates smallholders
- Improves techniques
- Increases yields and profits
- Improves social and educational standards

# Existing solutions

Transmission Type	Drawbacks	Benefits
Radio	People may not have radio Have to tune in at specific time Cannot choose what information to hear Not specific to region	Sometimes there are channels dedicated to agricultural announcements
Newspaper	Cost of purchasing newspaper Not specific to smallholding region Cannot choose what information they need	Reliable source
TV	People may not have TV Have to tune in at specific time Cannot choose what information to hear Not specific to region	Reliable source
SMS	Relies on people having phones with sms-capabilities May not support local dialects May only be Push or Pull	Can be tailored to user Low cost
Voice	Requires phone	Overcomes literacy barriers
Agricultural extension worker	Face to face personalised info	No accountability, may discriminate.

# Existing solutions

Transmission Type	Drawbacks	Benefits
Radio	People may not have radio Have to tune in at specific time Cannot choose what information to hear Not specific to region	Sometimes there are channels dedicated to agricultural announcements
Newspaper	Cost of purchasing newspaper Not specific to smallholding region Cannot choose what information they need	Reliable source
TV	People may not have TV Have to tune in at specific time Cannot choose what information to hear Not specific to region	Reliable source
SMS	Relies on people having phones with sms-capabilities May not support local dialects May only be Push or Pull	Can be tailored to user Low cost
Voice	Requires phone	Overcomes literacy barriers
Agricultural extension worker	Face to face personalised info	No accountability, may discriminate.

- Real time
  - Time critical messages
- Low cost
- Extensible solution

# Existing solutions

Transmission Type	Drawbacks	Benefits
Radio	People may not have radio Have to tune in at specific time Cannot choose what information to hear Not specific to region	Sometimes there are channels dedicated to agricultural announcements
Newspaper	Cost of purchasing newspaper Not specific to smallholding region Cannot choose what information they need	Reliable source
TV	People may not have TV Have to tune in at specific time Cannot choose what information to hear Not specific to region	Reliable source
SMS	Relies on people having phones with sms-capabilities May not support local dialects May only be Push or Pull	Can be tailored to user Low cost
Voice	Requires phone	Overcomes literacy barriers
Agricultural extension worker	Face to face personalised info	No accountability, may discriminate.

- Not push and pull
- Not natural language
  - Text 'R' for rice
- Doesn't take advantage of Agriculture 4.0
- All crops? Whole crop cycle?  
Globally?

# AES in practice - PxD



- 10 countries
- Total users in 2021: 5.7M

# AES in practice - PxD



**A/B tests generate insights for increasing user engagement.**

## FALL ARMYWORM MESSAGE DESIGN, PXD 2019

PxD has effectively increased farmer engagement with digital content by testing small tweaks in messaging through A/B trials. For example, in Kenya, PxD found that small, meaningful changes in the words and language used (for example, using urgent language) in messages can increase engagement by three percentage points.

## GUJARAT CALL LENGTH A/B TEST, PXD 2017

Results from A/B testing on customization features found that we could nudge farmers in India to listen to 28.8% more content if we gave them the option to choose between two short calls a week or a single, long call.

## ETHIOPIA, ROTATING IVR MENU OPTIONS, PXD 2018

Tweaking message framing – for example, by including a user's previous best quiz score in a messaged invitation to participate in a quiz about Fall Armyworm in Kenya – increased response rates by 11 percentage points over the control group. Rotating the ordering of agricultural content on a hotline audio menu in Ethiopia so that the most relevant option is always #1, resulted in a two-fold increase in farmers reaching the most timely content.

# AES in practice - PxD



**A/B tests generate insights for increasing user engagement.**

## FALL ARMYWORM MESSAGE DESIGN, PXD 2019

PxD has effectively increased farmer engagement with digital content by testing small tweaks in messaging through A/B trials. For example, in Kenya, PxD found that small, meaningful changes in the words and language used (for example, using urgent language) in messages can increase engagement by three percentage points.

## GUJARAT CALL LENGTH A/B TEST, PXD 2017

Results from A/B testing on customization features found that we could nudge farmers in India to listen to 28.8% more content if we gave them the option to choose between two short calls a week or a single, long call.

## ETHIOPIA, ROTATING IVR MENU OPTIONS, PXD 2018

Tweaking message framing – for example, by including a user's previous best quiz score in a messaged invitation to participate in a quiz about Fall Armyworm in Kenya increased response rates by 11 percentage points over the control group. Rotating the ordering of agricultural content on a hotline audio menu in Ethiopia so that the most relevant option is always #1, resulted in a two-fold increase in farmers reaching the most timely content.

# Research – machine learning based

	AgriBot	Intelligent Chatbot
Implementation	<ul style="list-style-type: none"><li>• Question answer dataset</li><li>• Sequence to sequence pattern</li><li>• Generative model</li></ul>	<ul style="list-style-type: none"><li>• Question answer dataset</li><li>• Cosine similarity</li><li>• API for weather and market prices.</li></ul>
Positives	Can respond to unseen input.	High accuracy. Incorporates real-time data.
Negatives	Limited to India. Only English. Not Natural Language. Not whole crop cycle.	Limited to India English and Hindi. No unseen input – refer to AEW

[1]

B. Arora, D. S. Chaudhary, M. Satsangi, M. Yadav, L. Singh, and P. S. Sudhish, ‘Agribot: A Natural Language Generative Neural Networks Engine for Agricultural Applications’, in *2020 International Conference on Contemporary Computing and Applications (IC3A)*, 2020, pp. 28–33.

[2]

R. Biswas and N. Goel, ‘Intelligent Chatbot Assistant in Agriculture Domain’, in *Agriculture-Centric Computation*, 2023, pp. 180–194.

# Research – machine learning based

AgriBot		Intelligent Chatbot
Implementation		<ul style="list-style-type: none"><li>• Question answer dataset</li><li>• Cosine similarity</li><li>• API for weather and market prices.</li></ul>
Positives		<p>High accuracy. Incorporates real-time data.</p>
Negatives		<p>Limited to India English and Hindi. No unseen input – refer to AEW</p>

[1]

B. Arora, D. S. Chaudhary, M. Satsangi, M. Yadav, L. Singh, and P. S. Sudhish, ‘Agribot: A Natural Language Generative Neural Networks Engine for Agricultural Applications’, in *2020 International Conference on Contemporary Computing and Applications (IC3A)*, 2020, pp. 28–33.

[2]

R. Biswas and N. Goel, ‘Intelligent Chatbot Assistant in Agriculture Domain’, in *Agriculture-Centric Computation*, 2023, pp. 180–194.

# Research – machine learning based

	AgriBot	Intelligent Chatbot
Implementation	<ul style="list-style-type: none"><li>• Question answer dataset</li><li>• Sequence to sequence pattern</li><li>• Generative model</li></ul>	<ul style="list-style-type: none"><li>• Question answer dataset</li><li>• Cosine similarity</li><li>• API for weather and market prices.</li></ul>
Positives	Can respond to unseen input.	High accuracy. Incorporates real-time data.
Negatives	Limited to India. Only English. Not Natural Language. Not whole crop cycle.	Limited to India English and Hindi. No unseen input – refer to AEW

[1]

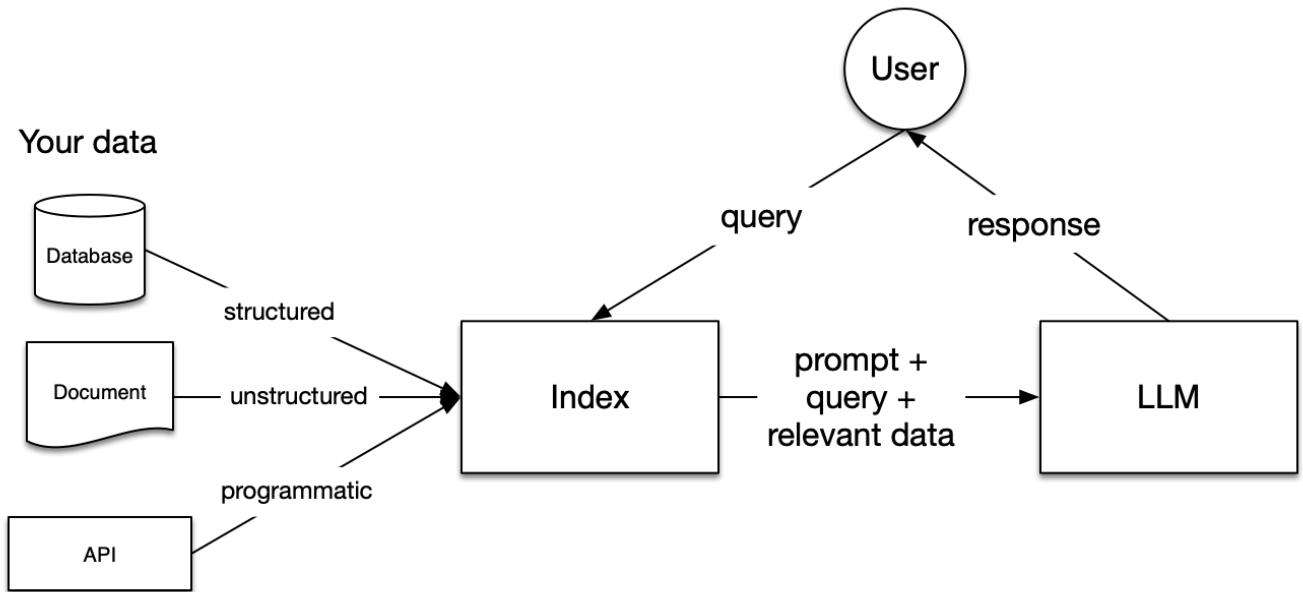
B. Arora, D. S. Chaudhary, M. Satsangi, M. Yadav, L. Singh, and P. S. Sudhish, ‘Agribot: A Natural Language Generative Neural Networks Engine for Agricultural Applications’, in *2020 International Conference on Contemporary Computing and Applications (IC3A)*, 2020, pp. 28–33.

[2]

R. Biswas and N. Goel, ‘Intelligent Chatbot Assistant in Agriculture Domain’, in *Agriculture-Centric Computation*, 2023, pp. 180–194.

# Terminology

What is a RAG?



- Retrieval Augmented Generation
- Context augmentation pattern.
- Retrieve data, and add it to a question as context

# Research - Large Language Models

- <10 studies done
- Most are specific eg ‘Leveraging Large Language Models for Extension Services in Nigeria’
- Only two are general
  - Both published after Progress Report

# Paper 1 Takeaways

- Recent study on using LLMs in AES systems
- Used OpenAI GPT and fine-tuned (manual process requiring experts)
- Found accuracy was not sufficient and suggested human integration

A. Tzachor *et al.*, “Large language models and agricultural extension services,” *Nat. Food*, vol. 4, no. 11, pp. 941–948, 2023, doi: 10.1038/s43016-023-00867-x

analytics-driven recommendation frameworks to provide data-based responses to questions such as ‘How much fertilizer should I apply to my cassava crop to realize a yield of crop-X tons per hectare? This response would ideally consider the farmer’s location (relevant for soil, weather and market conditions), but we are still far from such capabilities, which will require technology advances, large corpora of open data adhering to FAIR Principles<sup>33</sup>—findability, accessibility, interoperability and reusability—and improved human-machine interfaces at minimum.’

Providing such advances, GPT might also be used to synthesize data about farmers, including information about farm operations, agricultural input usage preferences and patterns, and environmental, economic and market conditions. By converting these data into easily understandable language—and always with humans involved in validating outputs—GPT could facilitate communicating improved personalized recommendations for growers that are appropriate to their specific needs, technical levels and circumstances. Incorporating the conversational history between the chatbot and each farmer can further enhance the chatbot’s ability to generate tailored recommendations, as it allows the system to continuously learn and adapt to the individual’s specific needs, preferences and context. With this added benefit, GPT-based applications can provide even more accurate and relevant guidance to farmers, fostering better decision-making and improved farm management practices. Chatbots serve as a convenient interface that can

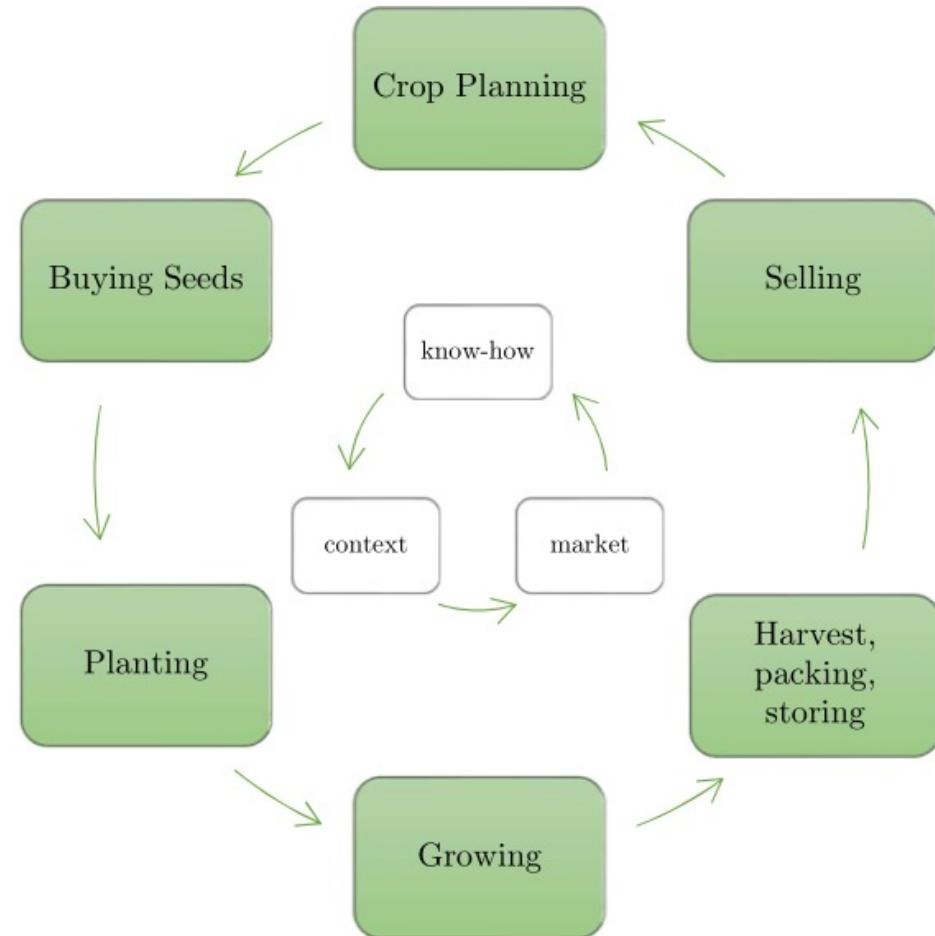
# Paper 2 Takeaways

Model	Fine-tuned	Accuracy	+RAG
Llama-2-chat 13B		75% ±2%	
Vicuna		79% ±2%	
GPT-4		80% ±4%	
Llama2 13B	✓	68% ±3%	77% ±2%
GPT-4	✓	81%±5%	86% ±2%

- OpenAI best LLM for base
- RAG and fine-tuning perform the same
  - RAG is better at being succinct
- Show some tailoring can be done (3 countries, Growing)

# Terminology

What is a crop cycle?



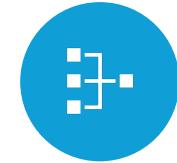
# Contents



Motivations



Research



System



Expert  
System



Adaptiveness



Other  
Achievements



Conclusions

# Objectives Summarised

- Natural Language Conversation
- Send notifications (temporal and event-driven)
- Whole crop cycle over most common cash crops
- Incorporate Agriculture 4.0
- Adaptive (quantitative – location; qualitative – literacy level)

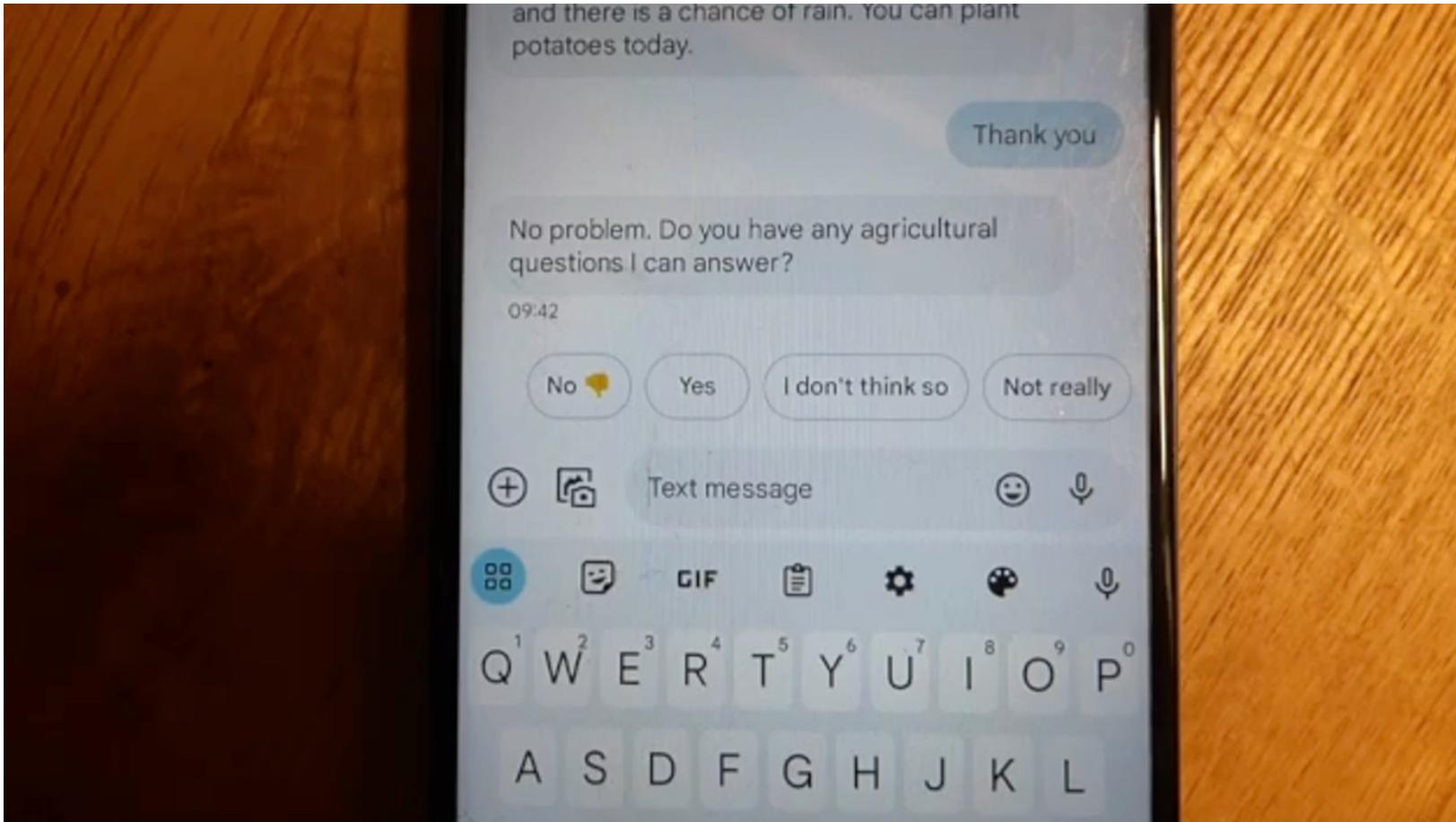
# Objectives Summarised

- Natural Language Conversation
- Send notifications (temporal and event-driven)
- Whole crop cycle over most common cash crops
- Incorporate Agriculture 4.0
- Adaptive (quantitative – location; qualitative – literacy level)



Not done within the context of agriculture

# Demonstration



and there is a chance of rain. You can plant potatoes today.

Thank you

No problem. Do you have any agricultural questions I can answer?

09:42

No 🤦

Yes

I don't think so

Not really



Text message



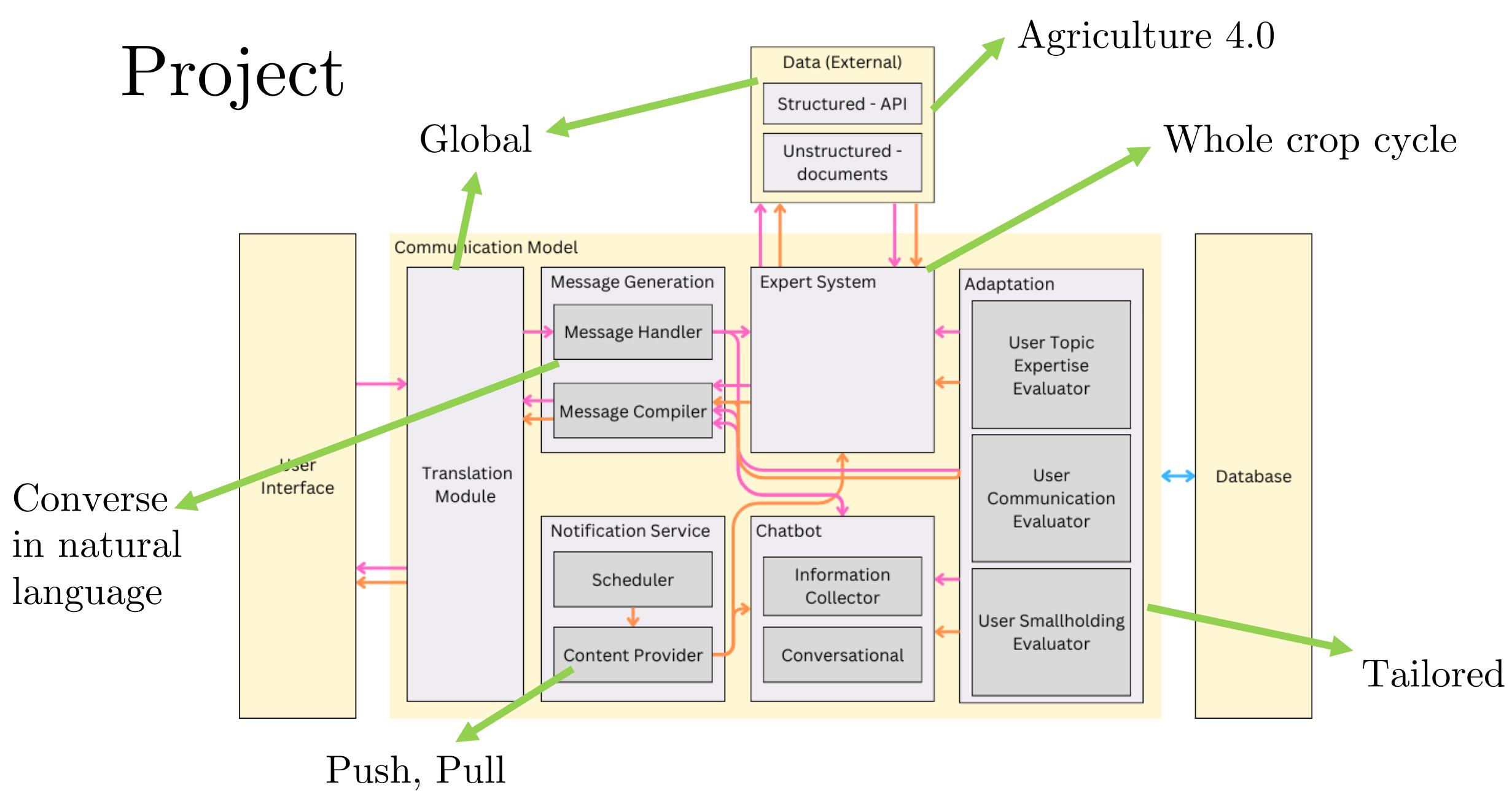
GIF



Q<sup>1</sup> W<sup>2</sup> E<sup>3</sup> R<sup>4</sup> T<sup>5</sup> Y<sup>6</sup> U<sup>7</sup> I<sup>8</sup> O<sup>9</sup> P<sup>0</sup>

A S D F G H J K L

# Project



# Technology Stack



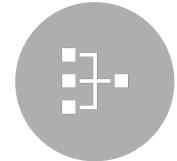
# Contents



Motivations



Research



System



Expert  
System



Adaptiveness



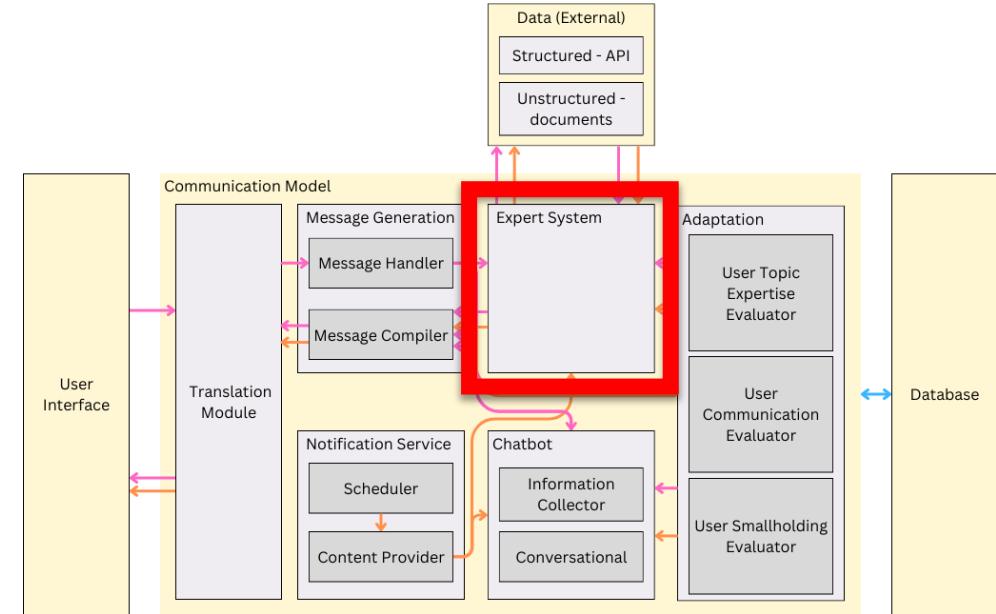
Other  
Achievements



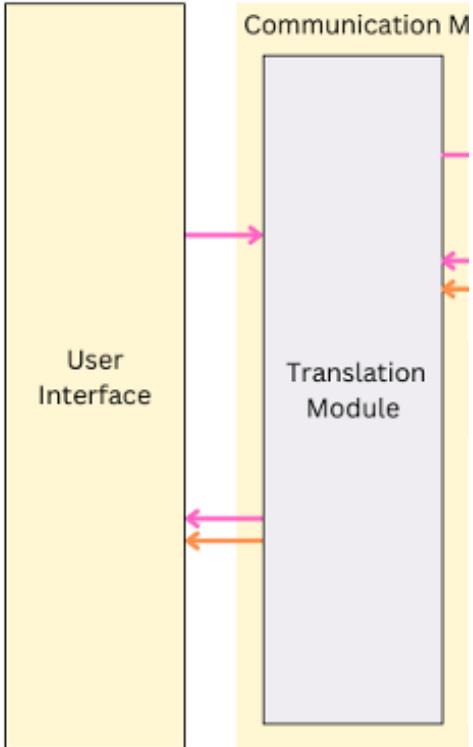
Conclusions

# Expert System – Initial thoughts

- Need to make a domain-specific LLM
- Can fine tune models, but
  - Expensive
  - Hard to keep up to date
  - No Observability
- Comparable accuracy with RAG



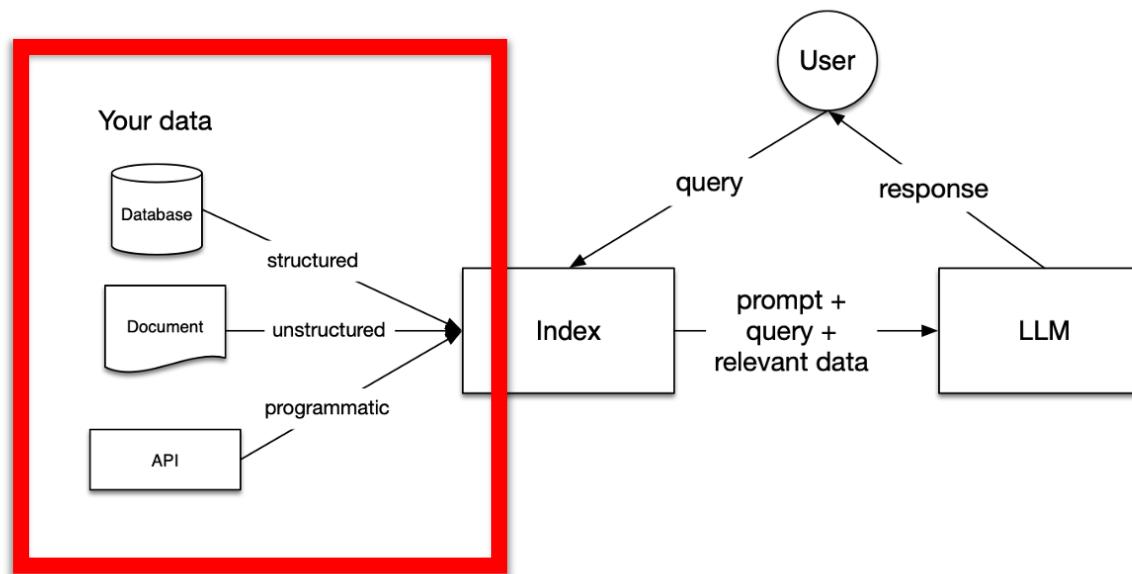
# Why Translate?



- Shown OpenAI ‘thinks’ in English
  - Despite multilingual capabilities
- More accurate responses
- Don’t have to worry about different languages in rest of system

# Benefits of RAG

- Allows for incorporating conversational data
- Context aware
- User can ask for clarifying information or give some themselves
- Takes structured and unstructured data – less difficulty

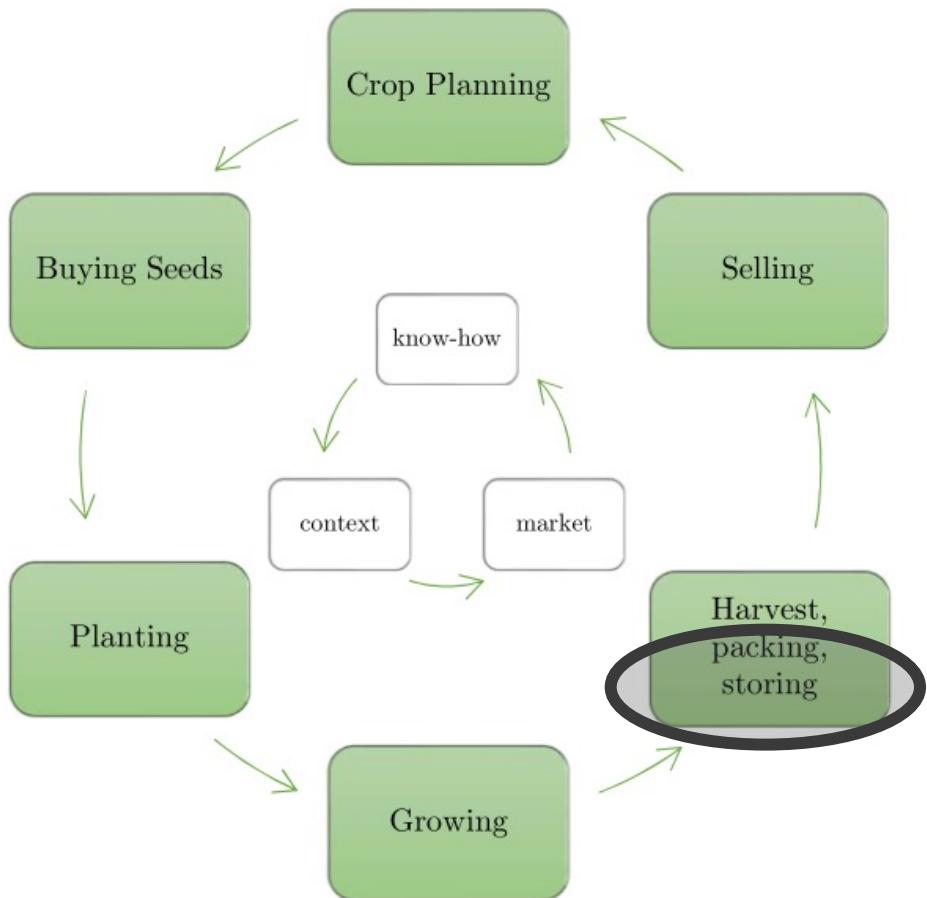


# Data Challenges

- No datasets or documents
  - Existing APIs provided only subset
- Should be extensible by Agriculture 4.0 – Soil, Plant, Weather, Climate location API

All scraped data will be published in own RESTful API libraries

# Expert System – Packing, storing



- No global hub of information on how each plant is packed and stored
- Divided plants into 26 groups as per IR-4 project
- Each group has different needs

# Expert System – Packing, storing (Limitations and Solutions)



- Map each to immediate (fresh) and long-term (dry) storing solutions
- Created document for llama-index

Packing and storing instructions for 25 crop groups.

#### Root and Tuber Veg:

Quick storing solution: Trim off greens

Dry storing solution: Allow some root cr

#### Leaves of Root and Tuber Veg. (Human Food or

Quick storing solution: Submerge the lea

Dry storing solution: Dehydrate leaves i

#### Bulb Veg:

Store in a cool, dry place with good vent

#### Leafy Veg:

Store in the refrigerator crisper drawer or in a plast

How can I store the last of my beetroots?  
And how should I pack them when I'm ready to sell?

You can store the last of your beetroots in a cold dry place with ample ventilation.

To pack them when you're ready to sell, trim off greens and pack the roots in perforated plastic bags or mesh bags.

# Expert System – Buying Seeds



- No global index of seed prices
- Take UK prices and use Purchasing Power Parity to estimate prices in other countries.

# Expert System – Buying Seeds



- Standard formulas to calculate seed rate (how much seed is needed)

How many wheat seeds should I buy? I want to cover half my smallholding

You need 134 grams of wheat seed.

# Expert System – Buying Seeds

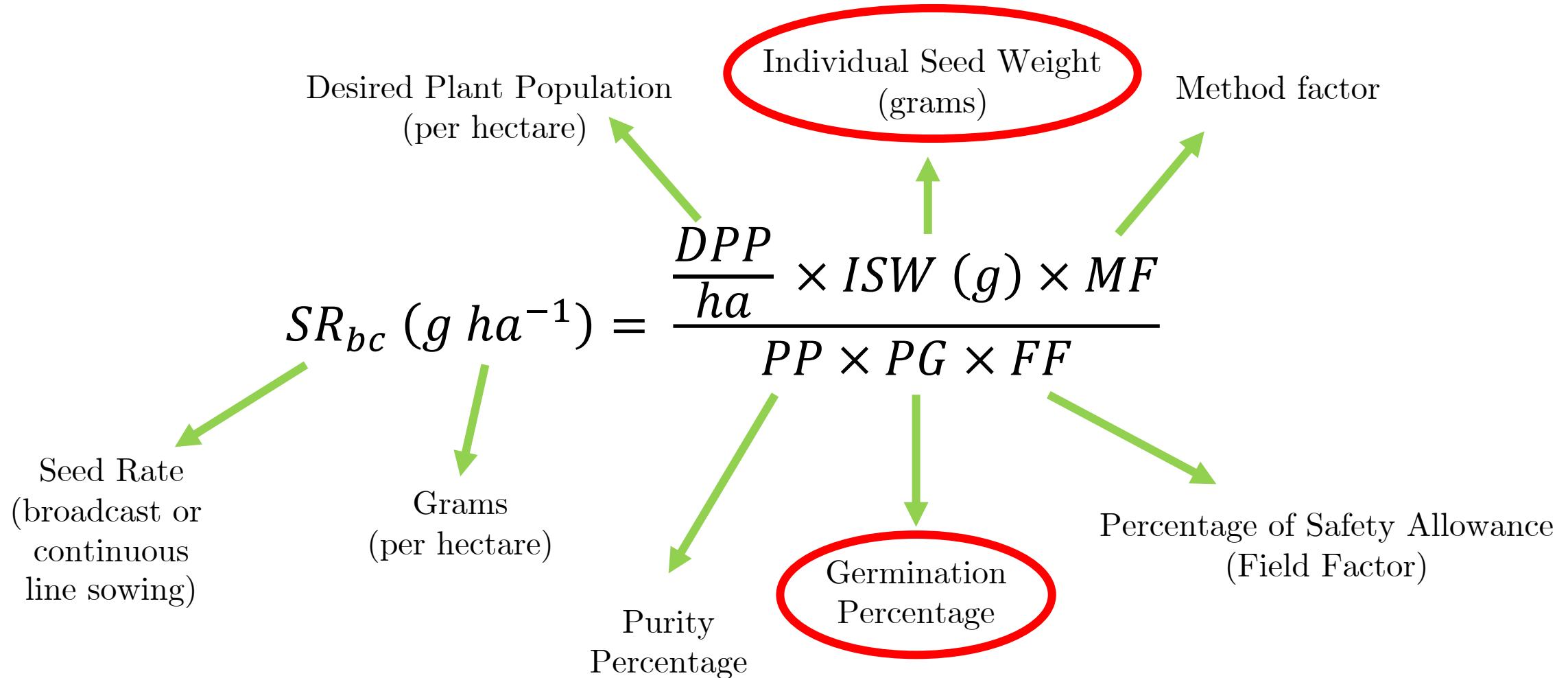
$$SR_{bc} (g \text{ } ha^{-1}) = \frac{\frac{DPP}{ha} \times ISW (g) \times MF}{PP \times PG \times FF}$$

The diagram illustrates the inputs to the seed rate calculation equation. It features a central equation with arrows pointing from various parameters to its components. The parameters are:

- Desired Plant Population (per hectare)
- Individual Seed Weight (grams)
- Method factor
- Percentage of Safety Allowance (Field Factor)
- Germination Percentage
- Purity Percentage
- Grams (per hectare)
- Seed Rate (broadcast or continuous line sowing)

Arrows point from the 'Desired Plant Population' and 'Individual Seed Weight' labels to the top terms of the numerator. Arrows point from 'Method factor', 'Percentage of Safety Allowance', 'Germination Percentage', 'Purity Percentage', 'Grams (per hectare)', and 'Seed Rate' to the bottom term of the denominator. Arrows also point from 'Desired Plant Population' and 'Grams (per hectare)' to the middle term of the numerator.

# Expert System – Buying Seeds



# Expert System – Buying Seeds



- Seed Information Database
- Run by SER and Kew
- No API or Dataset
- Request using their API Key
- Stored in Database
- RESTful API

```
GET /rest/v1/species_summary?select=*&or=%28has.  
Host: fxheguykvewpdeysvoh.supabase.co  
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac O  
Accept: */*  
Accept-Language: en-GB,en;q=0.5  
Accept-Encoding: gzip, deflate, br  
Referer: https://ser-sid.org/  
X-Client-Info: supabase-js/1.35.7  
apikey: eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.ey  
Authorization: Bearer eyJhbGciOiJIUzI1NiIsInR5c
```

# Expert System – Buying Seeds

$$SR_{bc} (g \text{ ha}^{-1}) = \frac{\frac{DPP}{ha} \times ISW (g) \times MF}{PP \times PG \times FF}$$

The diagram illustrates the factors influencing seed rate ( $SR_{bc}$ ). The formula is:

$$SR_{bc} (g \text{ ha}^{-1}) = \frac{\frac{DPP}{ha} \times ISW (g) \times MF}{PP \times PG \times FF}$$

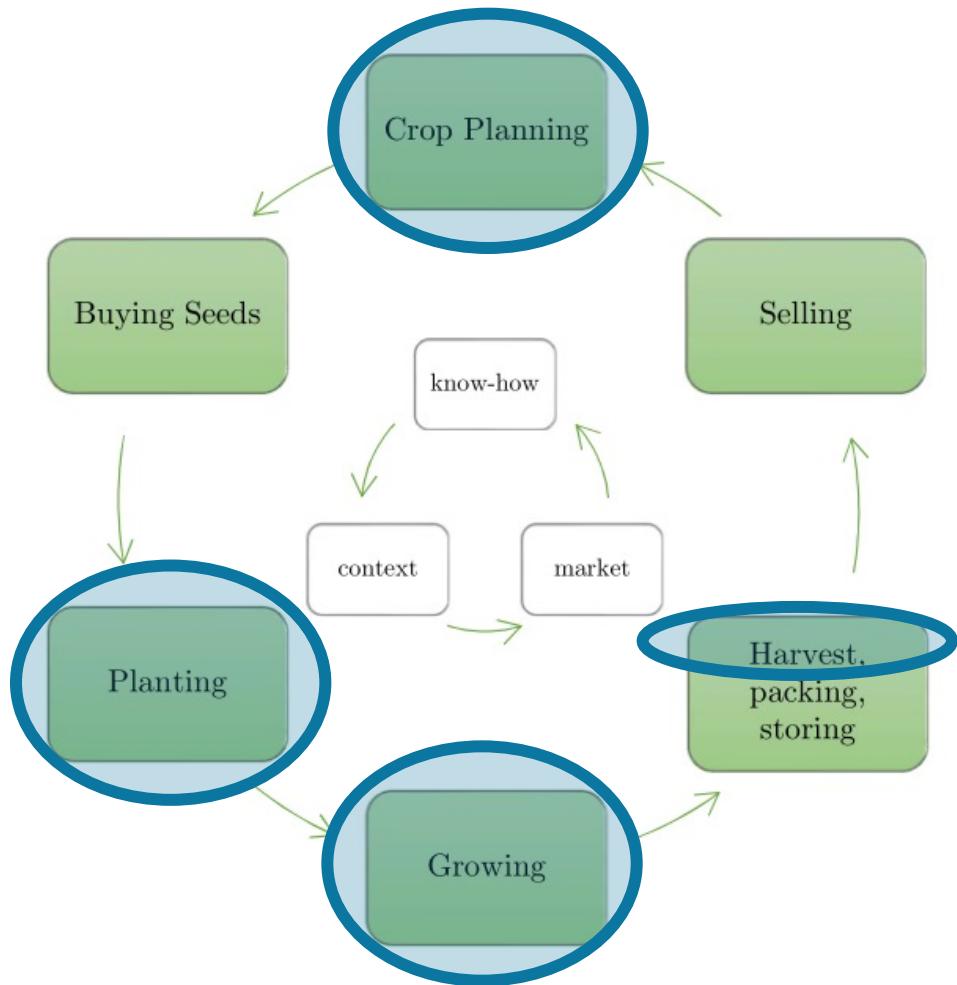
The factors are:

- Desired Plant Population (per hectare)
- Individual Seed Weight (grams)
- Method factor
- Seed Rate (broadcast or continuous line sowing)
- Grams (per hectare)
- Purity Percentage
- Germination Percentage
- Percentage of Safety Allowance (Field Factor)

Green arrows point from each factor to its corresponding term in the formula.

- Studies done that soil type and climate affects seed rate
- But no formulas produced
- Find optimal conditions for crop and linearly increase seed rate the further away we go (as more crops fail and crops produce less)

# Expert System – Planning, Growing, Harvesting



- Use Trefle API for general knowledge
- But many columns are NULL

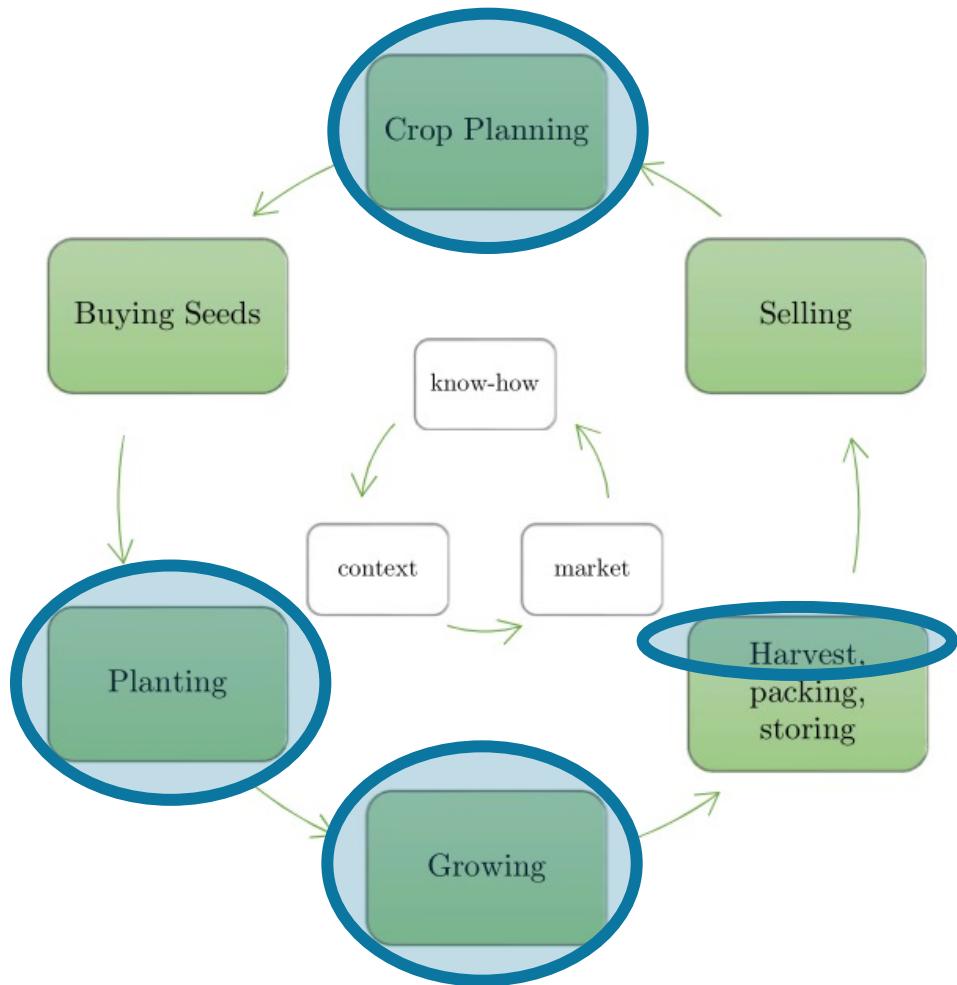
# Expert System – Planning, Growing, Harvesting

```
▼ specifications:  
  ligneous_type: null  
  growth_form: "Bunch"  
  growth_habit: "Graminoid"  
  growth_rate: "Rapid"  
  ▼ average_height:  
    cm: null  
  ▼ maximum_height:  
    cm: null  
    nitrogen_fixation: null  
    shape_and_orientation: null  
    toxicity: null
```

Wheat

```
▼ growth:  
  description: null  
  sowing: null  
  days_to_harvest: null  
  ▼ row_spacing:  
    cm: null  
  ▼ spread:  
    cm: null  
    ph_maximum: 8  
    ph_minimum: 5.5  
    light: 9  
    atmospheric_humidity: 3  
    growth_months: null  
    bloom_months: null  
    fruit_months: null  
  ▼ minimum_precipitation:  
    mm: null  
  ▼ maximum_precipitation:  
    mm: null  
  ▼ minimum_root_depth:  
    cm: null  
  ▼ minimum_temperature:  
    deg_f: null  
    deg_c: null  
  ▼ maximum_temperature:  
    deg_f: null  
    deg_c: null  
    soil_nutrients: 7  
    soil_salinity: 0  
    soil_texture: null  
    soil_humidity: null
```

# Expert System – Planning, Growing, Harvesting



- Use Trefle API for general knowledge
- But many columns are NULL
- So amalgamate all sub-species

```
▲ Jacqueline Dobreva-Skevington
final inline fun <reified T> defaultNoArgConstructor(): T{
    return T::class.constructors.first { it.parameters.isEmpty() }.call()
}

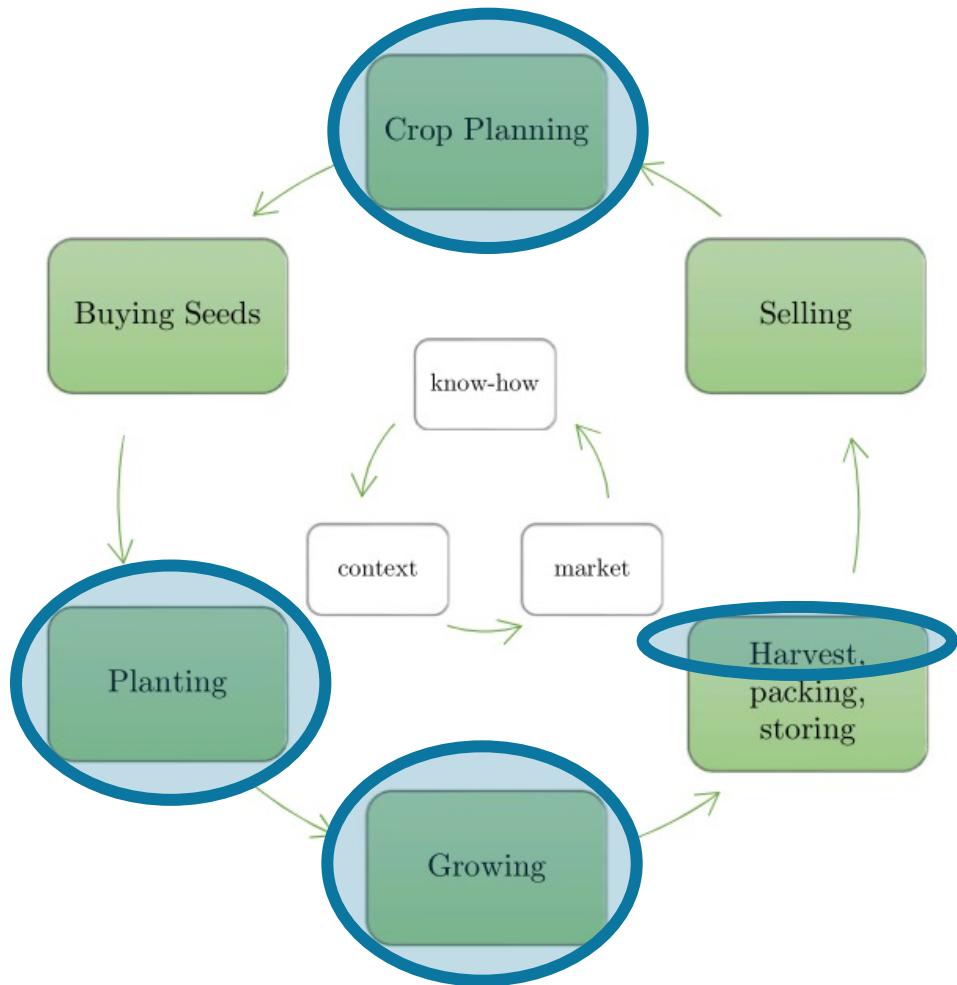
▲ Jacqueline Dobreva-Skevington
final inline fun <reified T> List<T>.coalesce(): T{
    val newObj = defaultNoArgConstructor<T>()!!
    newObj::class.memberProperties.mapNotNull { it as? KMutableProperty<*> }.forEach {kProp->
        val firstNotNullOrNull = this.firstNotNullOrNull { kProp.getter.call(it) }
        kProp.setter.call(newObj, firstNotNullOrNull)
    }
    return newObj
}
```

# Expert System – Planning, Growing, Harvesting

```
{  
    "days_to_harvest" : null,  
    "description" : null,  
    "sowing" : "Direct seed, thin seedlings to 15cm",  
    "ph_maximum" : 8.0,  
    "ph_minimum" : 5.5,  
    "light" : 9,  
    "atmospheric_humidity" : 3,  
    "growth_months" : null,  
    "bloom_months" : [ "jun", "jul", "aug" ],  
    "fruit_months" : null,  
    "spacing_between_rows" : null,  
    "average_spreading" : null,  
    "min_precipitation_per_year" : null,  
    "max_precipitation_per_year" : null,  
    "min_soil_depth" : null,  
    "min_temperature" : null,  
    "max_temperature" : null,  
    "soil_nutriments" : 7,  
    "soil_salinity" : 0,  
    "soil_texture" : null,  
    "soil_humidity" : null  
}
```

```
▼ growth:  
  description: null  
  sowing: null  
  days_to_harvest: null  
  ▼ row_spacing:  
    cm: null  
  ▼ spread:  
    cm: null  
    ph_maximum: 8  
    ph_minimum: 5.5  
    light: 9  
    atmospheric_humidity: 3  
    growth_months: null  
    bloom_months: null  
    fruit_months: null  
  ▼ minimum_precipitation:  
    mm: null  
  ▼ maximum_precipitation:  
    mm: null  
  ▼ minimum_root_depth:  
    cm: null  
  ▼ minimum_temperature:  
    deg_f: null  
    deg_c: null  
  ▼ maximum_temperature:  
    deg_f: null  
    deg_c: null  
  soil_nutriments: 7  
  soil_salinity: 0  
  soil_texture: null  
  soil_humidity: null
```

# Expert System – Planning, Growing, Harvesting



- Use Trefle API for general knowledge
- But many columns are NULL
- So amalgamate all sub-species
- Use SID
- And use ECOCROP (FAOSTAT)
- No API or csv export

# Expert System – Planning, Growing, Harvesting

Ecology							
	Optimal		Absolute				
	Min	Max	Min	Max	Soil depth	shallow (20-50 cm)	shallow (20-50 cm)
Temperat. requir.	20	30	12	35	Soil texture	heavy, medium, light, organic	heavy, medium, light
Rainfall (annual)	600	1200	300	2500	Soil fertility	high	moderate

Console	Inspector	Debugger	Network	Style Editor	Performance	Memory	Storage	Accessibility	Application	...	Raw		
Filter URLs													
is	Method	Domain	File	Initiator	Type	Transferred	Size	Headers	Cookies	Request	Response	Timings	Security
GET		nupcon.arcgi...	iauencn10.json	<a href="#">p-uearecaaa.ent...</a>	json	cached	78 B	HTML	All	HTML	Raw		
GET		hubcdn.arcgi...	x24.json	<a href="#">p-0eaef5aa.ent...</a>	json	cached	132 B						
GET		hubcdn.arcgi...	messages_en.json	<a href="#">p-8cac1862.js...</a>	json	cached	27 B						
GET		hqfa0.maps....	data	<a href="#">FaviconLoader...</a>	png	cached	11.89 ...						
GET		ago-item-sto...	GAEZ_Logo.png?X-Amz-Security-Token=IQoJbG...	<a href="#">FaviconLoader...</a>	png	cached	11.89 ...						
GET		ago-item-sto...	fao-2log2.jpg?X-Amz-Security-Token=IQoJb3Jp...	<a href="#">vendor-d5292a...</a>	jpeg	195.40 kB	194.7...						
POST		pinpoint-fips...	events	<a href="#">chunk.51741.76...</a>	json	629 B	298 B						
GET		www.arcgis...	d482830372b9413c84d3c004f56e0889?f=json	<a href="#">chunk.51741.76...</a>	json	1.13 kB (raced)	1.22 kB						
GET		www.arcgis...	data?f=json	<a href="#">chunk.51741.76...</a>	json	2.69 kB (raced)	7.53 kB						
GET		www.arcgis...	d482830372b9413c84d3c004f56e0889?f=json	<a href="#">chunk.51741.76...</a>	json	1.13 kB (raced)	1.22 kB						
GET		www.arcgis...	data?f=json	<a href="#">chunk.51741.76...</a>	json	2.69 kB (raced)	7.53 kB						
GET		hubcdn.arcgi...	arcgis-shareable-card.18n.en.json	<a href="#">chunk.52930.1...</a>	json	103 B (raced)	131 B						
GET		ecocrop.revi...	cropFindForm	<a href="#">vendor-d5292a...</a>	html	4.79 kB	4.54 kB						
GET		hqfa0.maps....	data	<a href="#">FaviconLoader...</a>	png	cached	11.89 ...						
GET		ago-item-sto...	GAEZ_Logo.png?X-Amz-Security-Token=IQoJbG...	<a href="#">FaviconLoader...</a>	png	cached	11.89 ...						
GET		ecocrop.revi...	cropList?name=a&relation=beginsWith	subdocument	html	172.93 kB	172.6...						
OPTI...		pinpoint-fips...	events	fetch	json	465 B	0 B						
POST		pinpoint-fips...	events	<a href="#">chunk.51741.76...</a>	json	550 B	219 B						
GET		ecocrop.revi...	cropView?id=289	<a href="#">cropList?7...</a>	html	5.58 kB	5.34 kB						
GET		ecocrop.revi...	289.jpg	img	jpeg	14.67 kB (raced)	14.67 ...						
GET		ecocrop.revi...	dataSheet?id=289	subdocument	html	6.28 kB	6.02 kB						

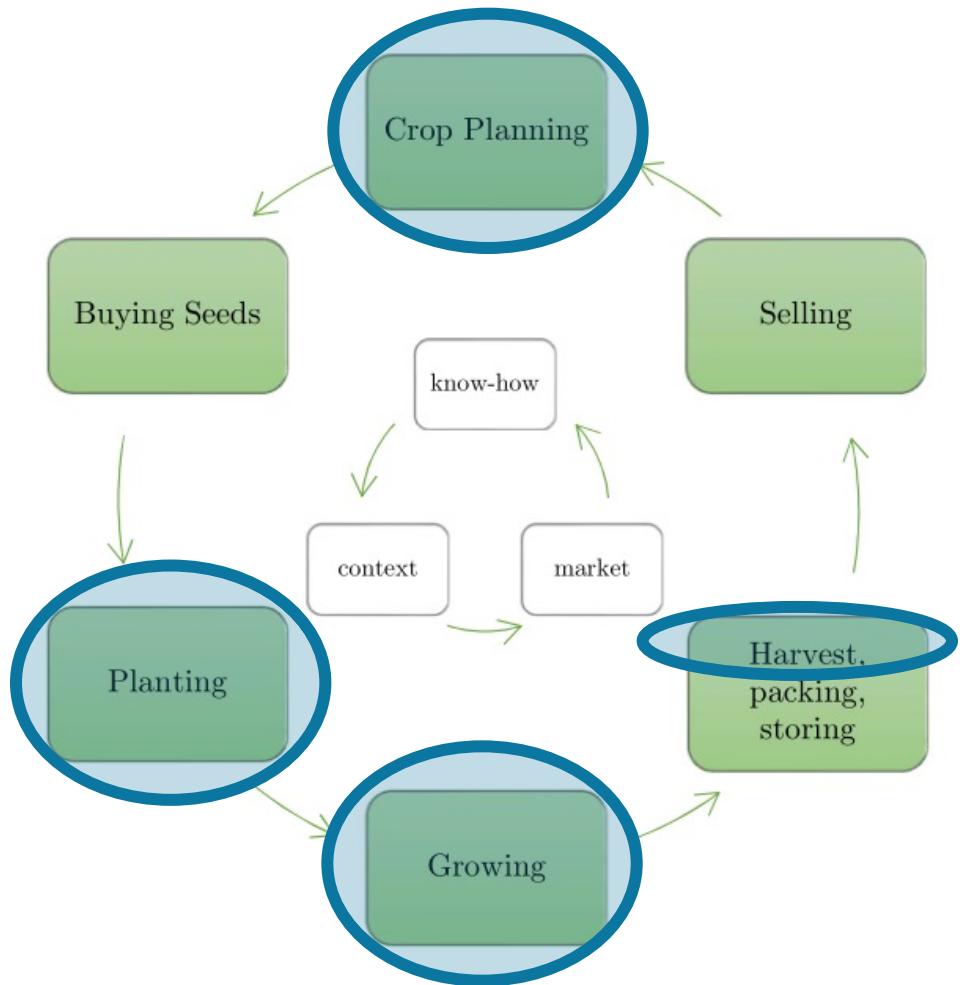
81 requests 1.17 MB / 654.94 kB transferred. Finish: 1.04 min DOMContentLoaded: 716 ms Load: 739 ms

# JSOUP

**Abelmoschus esculentus**

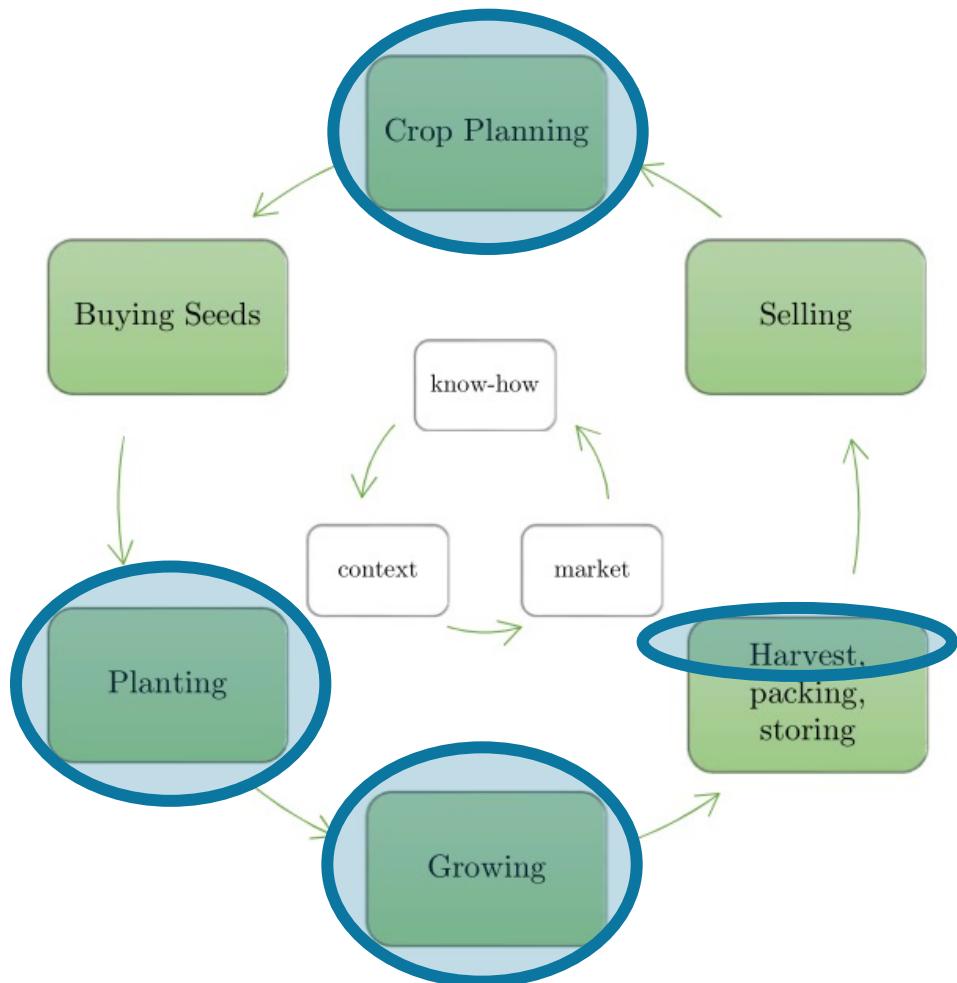
Description							
Life form	herb	Physiology	single stem				
Habit	erect	Category	vegetables				
Life span	annual	Plant attributes	grown on large scale				
Ecology							
Optimal	Absolute	Optimal	Absolute				
Min	Max	Min	Max				
Temperat. requir.	20	30	12	35	Soil depth	shallow (20-50 cm)	shallow (20-50 cm)
Rainfall (annual)	600	1200	300	2500	Soil texture	heavy, medium, light, organic	heavy, medium, light
Latitude	-	-	35	40	Soil Al. tox		
Altitude	---	---	-	1000	Soil salinity	low (<4 dS/m)	low (<4 dS/m)
Soil pH	5.5	7	4.5	8.7	Soil drainage	well (dry spells)	well (dry spells)
Light intensity	clear skies	clear skies	very bright	cloudy skies			
Climate zone	tropical wet & dry (Aw), tropical wet (Ar), steppe or semiarid (Bs), subtropical humid (Cf), subtropical dry summer (Cs), subtropical dry winter (Cw)				Photoperiod	short day (<12 hours)	
Killing temp. during rest	no input				Killing temp. early growth	no input	

# Expert System – Planning, Growing, Harvesting

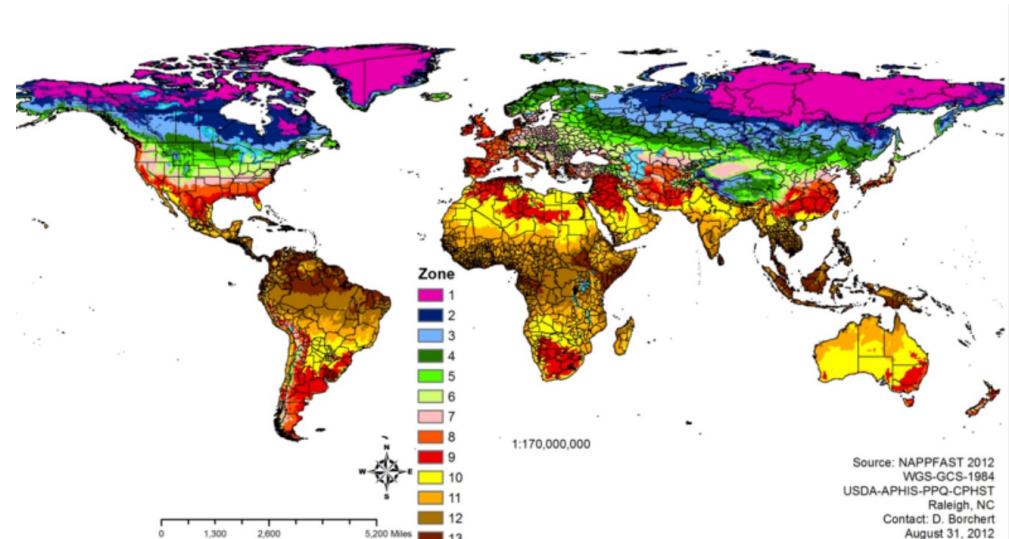


- Tailoring (A 4.0)
- Wheat sown in Wales may have very different requirements from Algeria

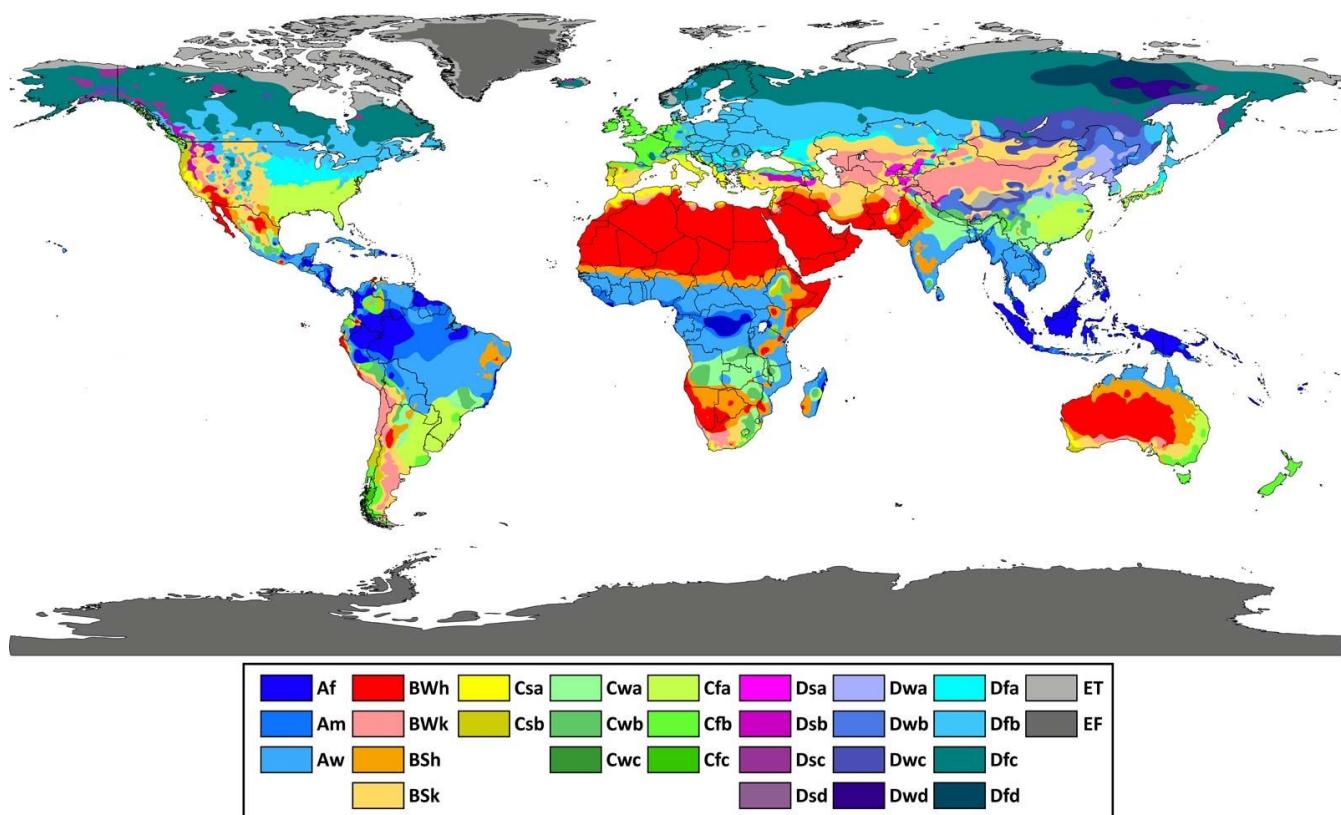
# Expert System – Planning, Growing, Harvesting



- World is divided into zones
- Standard is ‘hardiness’ zones
- But only based on temperature



# Expert System – Planning, Growing, Harvesting



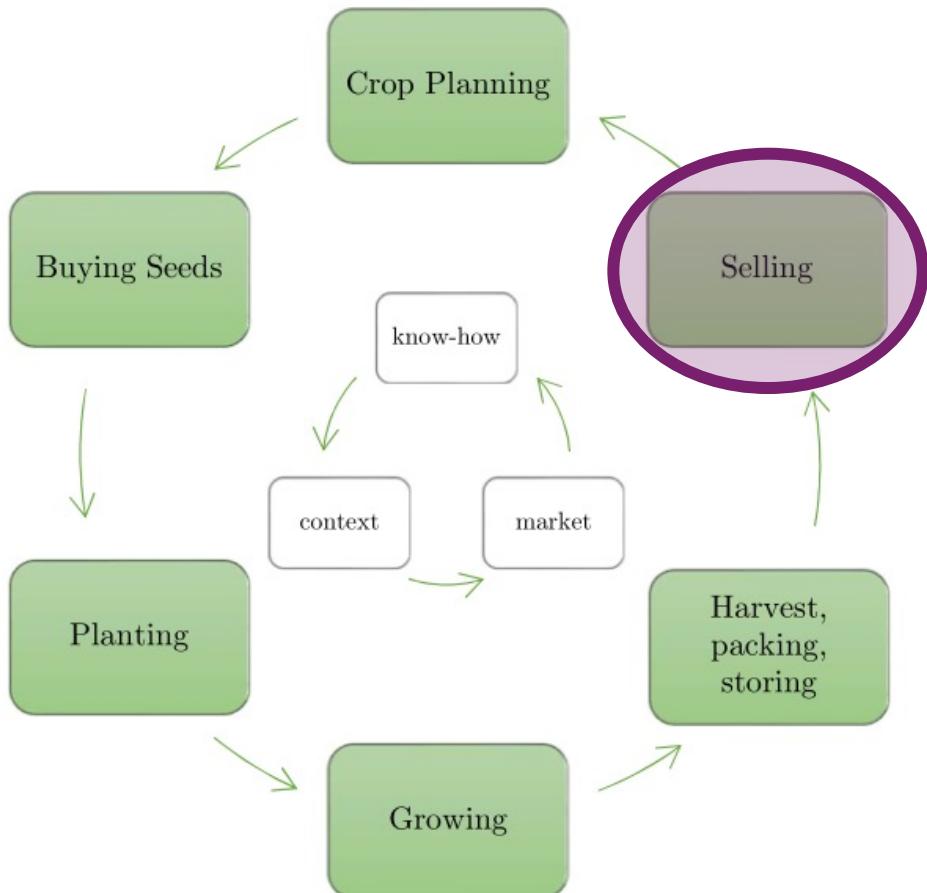
- Use Köppen climate zones
  - ECOCROP
- Each zone has own growing conditions
- Soil further affects these (eg clay soil takes longer to warm up but retains heat for longer)

Should I plant aubergine now?

The absolute minimum temperature for aubergine is 9°C. Therefore it is not advisable to plant aubergines now, as temperatures may drop below this threshold.

It's better to wait until June when temperatures are less likely to fall below 9°C.

# Expert System – Selling



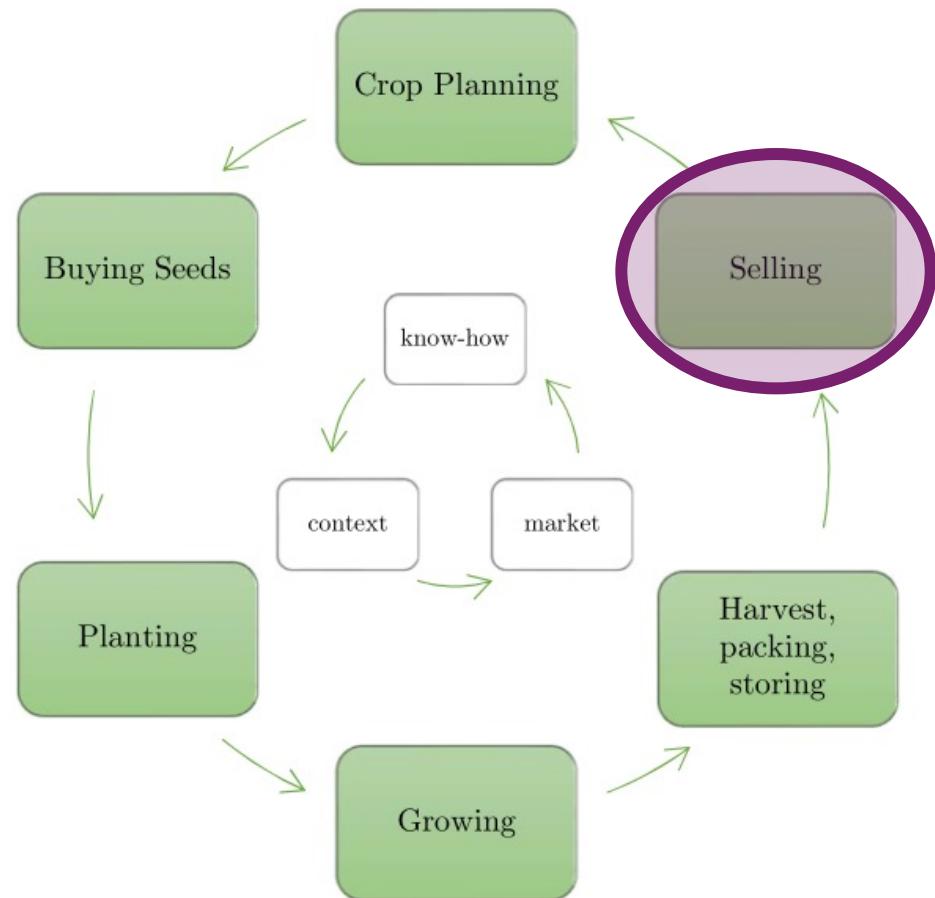
- Use FAOSTAT producer prices data (csv)
- Suggest selling when prices will be high
  - Buying Seeds?

Should I sell the rest of my potatoes?

As your smallholding is in the UK, if you sell your potatoes now you can fetch £244.56 per tonne.

However, if you wait for three months, the price might increase to £247.02 per tonne.

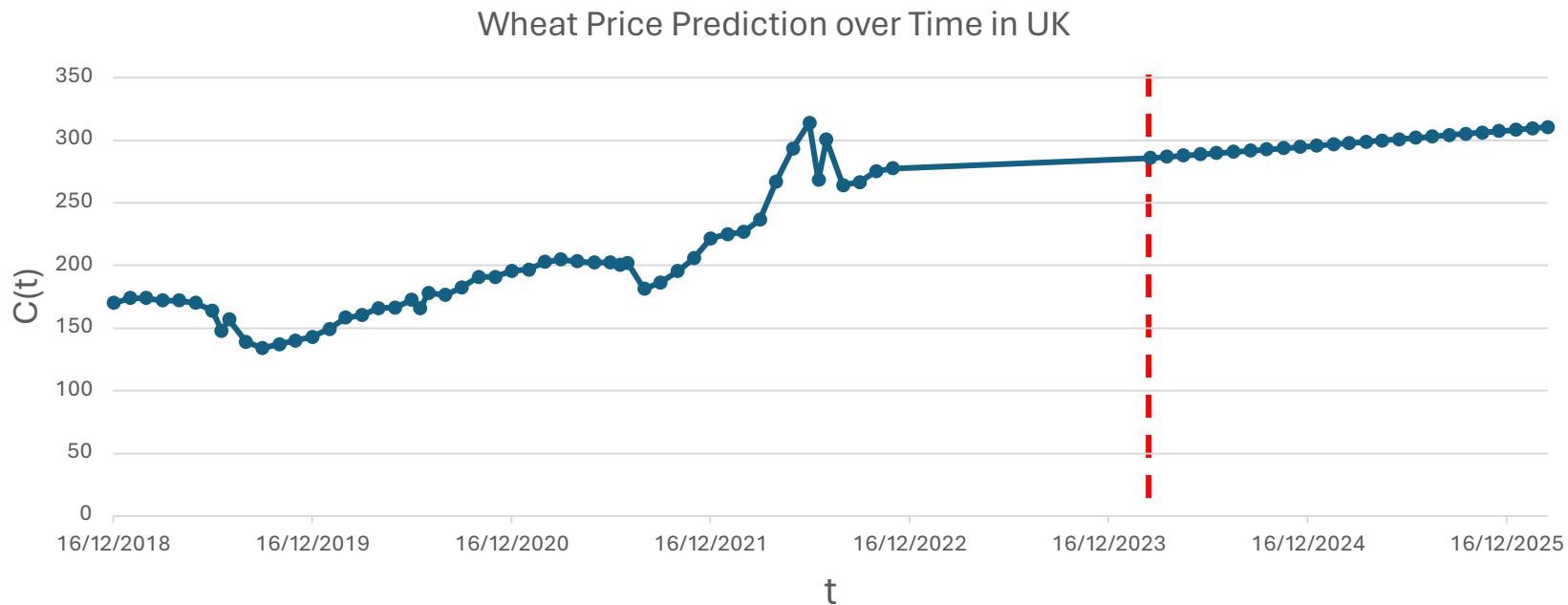
# Expert System – Selling (limitations and solutions)



- Farmers see little profit (0.03%) when prices are high
- Fluctuating demand
  - Not instant product
- Predict future price to inform crop planning and buying seed stage



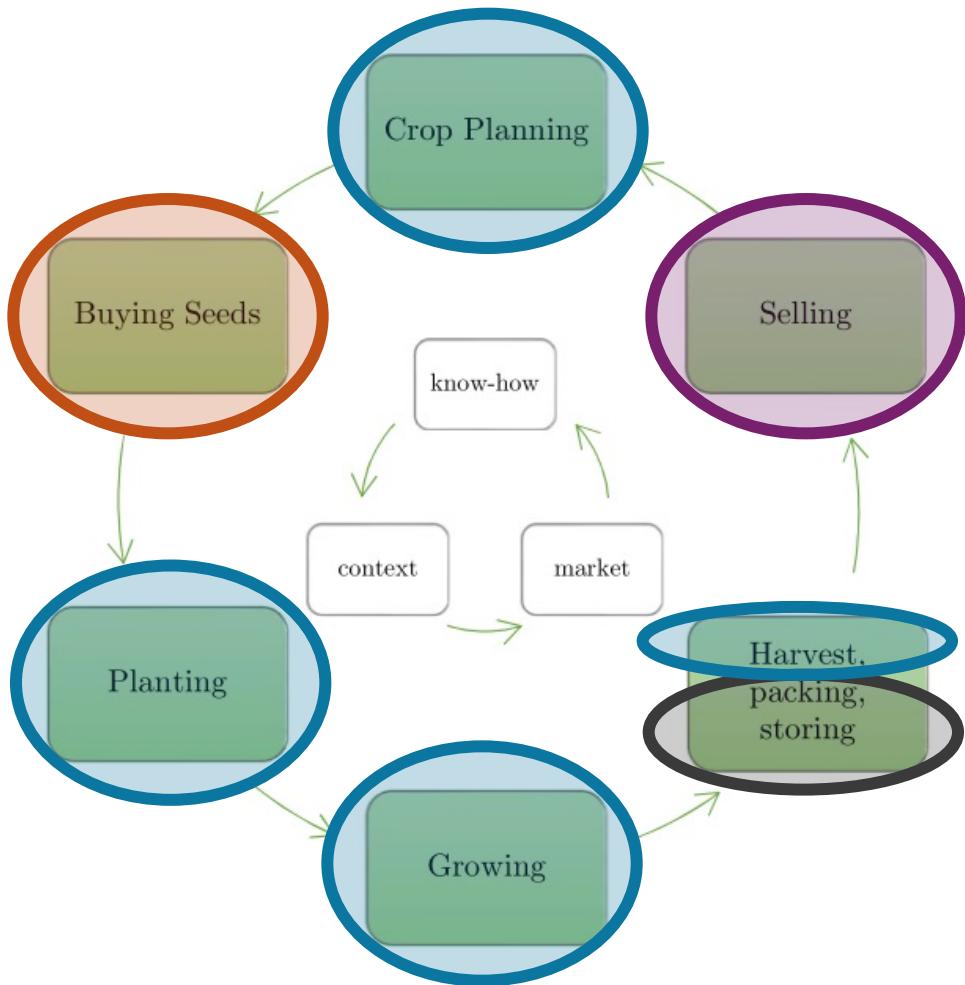
# Expert System – Selling (predictions)



Quadratically  
weighted linear  
average (weighted  
towards more recent  
data)  
Tunable ‘k’

$$C(t_p) - C(t_i) = \frac{\sum_{i=1}^n \left( \frac{C(t_0) - C(t_i)}{t_0 - t_i} * (t_p - t_i)^{-k} \right)}{\sum_{i=1}^n (t_p - t_i)^{-k}}$$

# Expert System – Tying it together



- Data available to llama-index
- ‘chatbot’ mode provides conversation context
- Multi-modal extensible approach

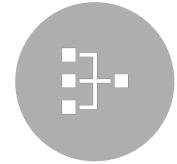
# Contents



Motivations



Research



System



Expert  
System



Adaptiveness

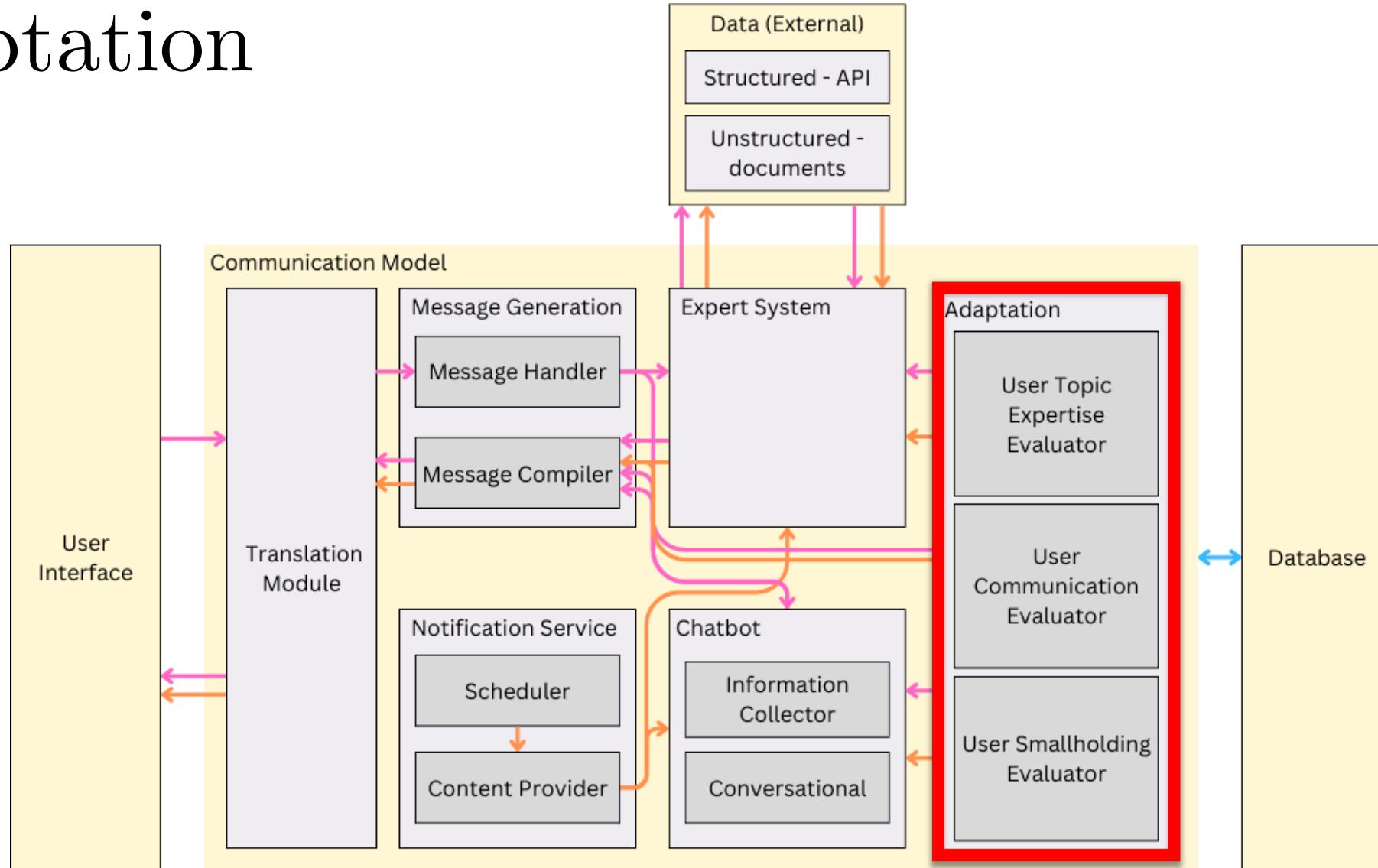


Other  
Achievements

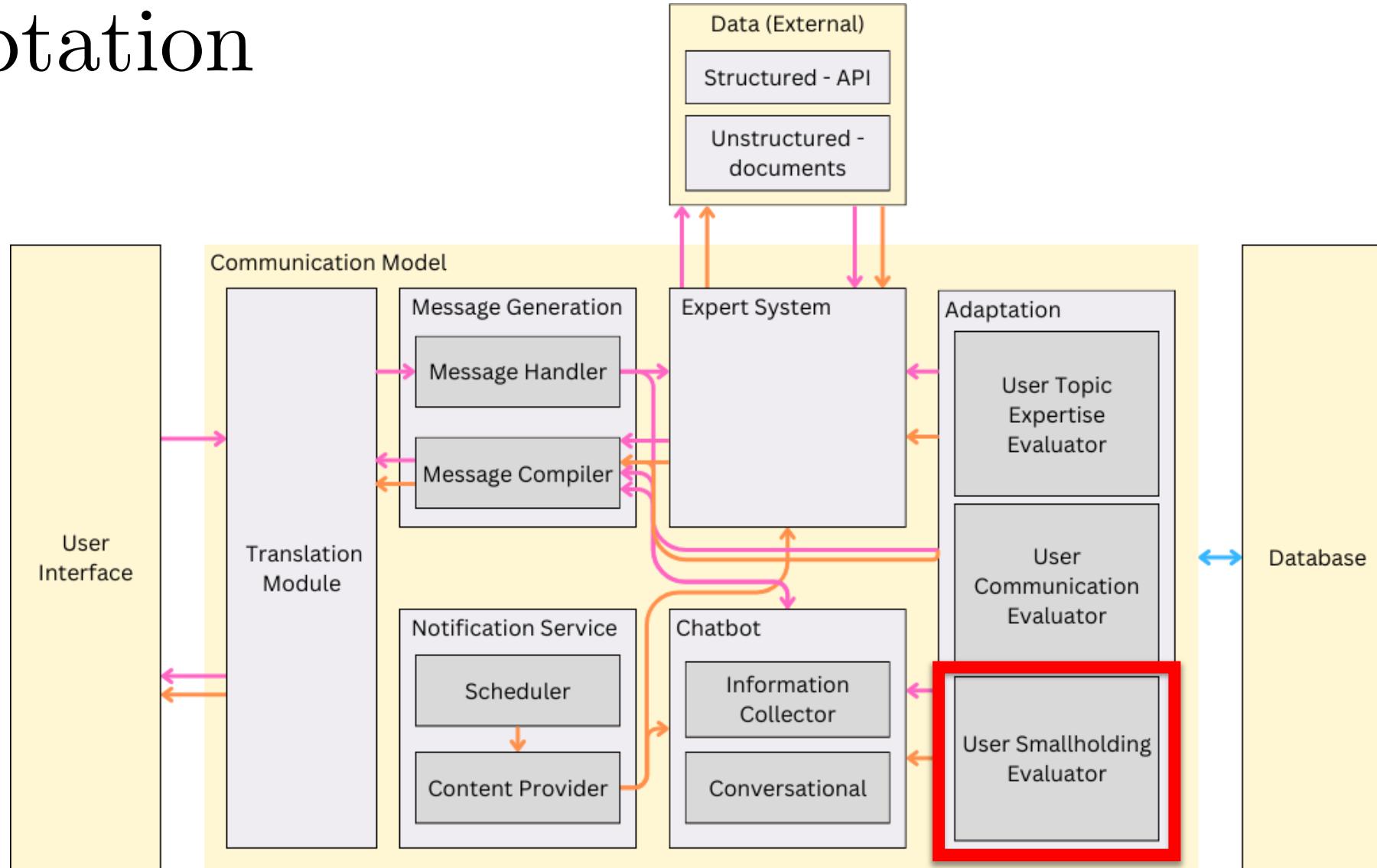


Conclusions

# Adaptation



# Adaptation



# Terminology

What is NER?

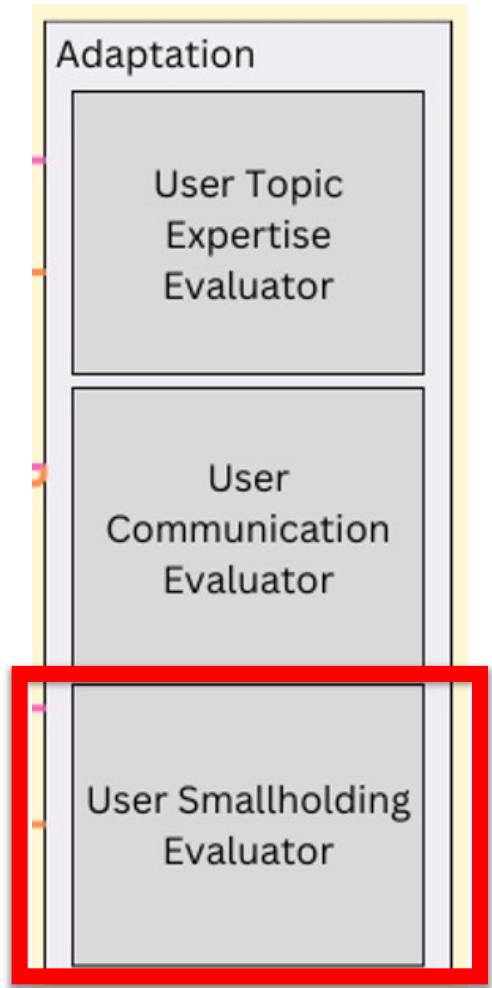
- Named Entity Recognition
- NLP technique
- Extracts entity from textual data into predefined categories

# Adaptation - Smallholding

- Libraries such as Spacy and NLTK
- Irrelevant categories
  - People
  - Organisations
  - Locations
  - Products
  - Events

*My smallholding in Swansea, Wales is 5 acres and I grow wheat. My name is Jacqueline.*

```
{<UserDetails.SMALLHOLDING_SIZE: 'SMALLHOLDING_SIZE': '5 acres',
<UserDetails.MAIN_CROP: 'MAIN_CROP': 'Wheat',
<UserDetails.NAME: 'NAME': 'Jacqueline',
<UserDetails.LOCATION_COUNTRY: 'LOCATION_COUNTRY': 'Wales',
<UserDetails.LOCATION_CITY: 'LOCATION_CITY': 'Swansea',}
```



# Adaptation - Smallholding

The screenshot shows a GAI (Generative AI) interface for adapting a language model to smallholding data. On the left, there's a sidebar with settings: 'Text Separator' (New Line), 'Annotation Precision' (Word Level), 'Tagging Progress (1/80)', and a 'Keyboard Shortcuts' toggle. The main area displays a sentence with tokens highlighted by colored boxes and numbered tags (1-5):

Hey there , I 'm Max Rodriguez NAME (1), based in Chicago LOCATION\_CITY (2), USA. LOCATION\_COUNTRY (3) My smallholding spans 2 hectares SMALLHOLDING\_SIZE (4)

Below the sentence, a question is shown: " , and I'm curious : How do I improve soil fertility for my maize and soybeans ?"

At the top right, there are buttons for 'NEW TAG' and 'EDIT TAGS'.

- No dataset so GAI generated and self tagged
- 80/20 split
- Trained for accuracy (larger language model) rather than efficiency

# Adaptation - Smallholding

- 96% accuracy (as using model-best)

```
===== Initializing pipeline =====
✓ Initialized pipeline

===== Training pipeline =====
ℹ Pipeline: ['tok2vec', 'ner']
ℹ Initial learn rate: 0.001

E   #      LOSS TOK2VEC  LOSS NER  ENTS_F  ENTS_P  ENTS_R  SCORE
--  --  -----  -----  -----  -----  -----  -----
  0    0       0.00     51.59    0.00     0.00     0.00     0.00
  7   200      88.29   1551.06   95.11    93.86    96.40     0.95
 16   400      14.89   173.54   94.55    95.41    93.69     0.95
 27   600      19.34   127.03   95.11    93.86    96.40     0.95
 41   800      22.85   123.99   96.00    94.74    97.30     0.96
 58  1000      20.08   113.59   95.54    94.69    96.40     0.96
 80  1200      53.17   160.20   95.07    94.64    95.50     0.95
 106 1400      63.54   177.88   95.11    93.86    96.40     0.95
 138 1600      59.06   187.34   93.75    92.92    94.59     0.94
 177 1800      10.54   119.49   95.07    94.64    95.50     0.95
 224 2000      10.76   132.02   93.75    92.92    94.59     0.94
 282 2200       5.13   146.05   92.44    91.23    93.69     0.92
 348 2400       8.77   171.10   92.92    91.30    94.59     0.93

✓ Saved pipeline to output directory
Data/model-best
```

```
# Import the spaCy library
import spacy
import sys

# Load the trained spaCy NER model from the specified path
nlp = spacy.load('Data/model-best')

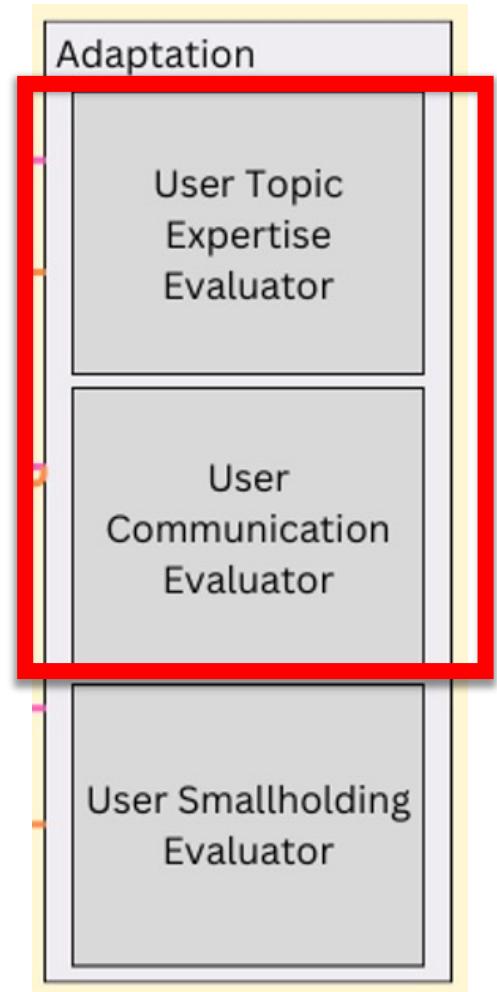
# Define the file name
fname = 'Data/test.txt'

# Open the PDF document using PyMuPDF (fitz)
with open("Data/test.txt", 'r') as file:
    text = file.read()
    doc = nlp(text)
    ents = doc.ents
    print([e.text for e in ents if e.label_ == "SMALLHOLDING_SIZE"])
    print([e.text for e in ents if e.label_ == "NAME"])
    print([e.text for e in ents if e.label_ == "LOCATION_CITY"])
    print([e.text for e in ents if e.label_ == "LOCATION_COUNTRY"])
    print([e.text for e in ents if e.label_ == "MAIN_CROP"])

ner-test ×
:
/usr/local/bin/python3.12 /Users/jacqueline.dobreva-skevington_t4c/Documents/Uni/AES-Q
['3 acres', '6-acre', '5 hectares', '4-acre', '2-hectare', '8 acres', '7-acre', '3-hec
['Daniel Wilson', 'Emily Davis', 'Sophia Smith', 'Liam Johnson', 'Olivia Brown', 'Etha
['Chicago', 'Sydney', 'London', 'Toronto', 'Auckland', 'Buenos Aires', 'Los Angeles',
['USA.', 'Australia.', 'UK.', 'Canada.', 'New Zealand.', 'Argentina.', 'USA.', 'Austra
['tomatoes', 'peppers', 'wheat', 'barley', 'oats', 'sunflowers', 'rapeseed', 'fruits',
```

# User suggestibility

- Sometimes smallholders don't implement suggestions
  - PxD
- Information is only useful if it has an affect
- Two proof of concepts
  - Tailoring on Knowledge level
  - Tailoring on User Characteristics

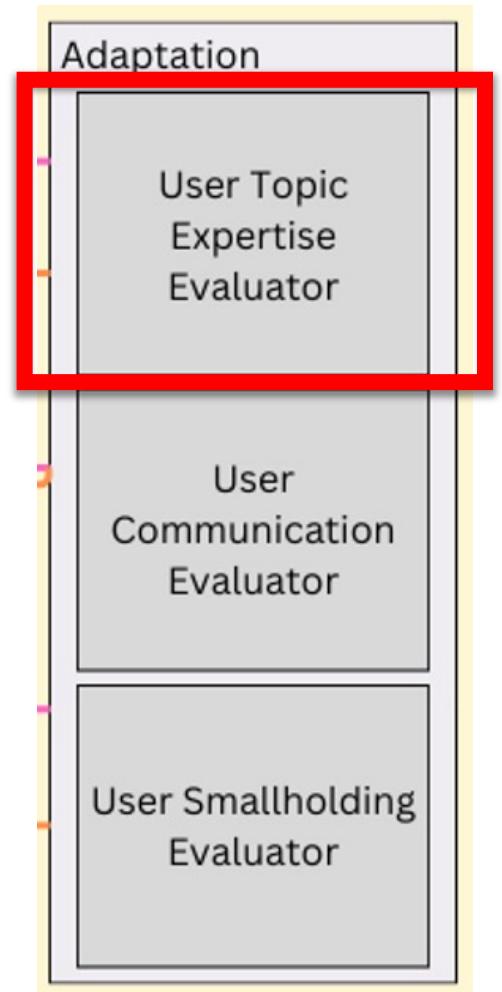


# Short Text Authorship Characteristic Problem

- Studies on identifying author of text
  - Characteristics
  - Matching
- Usually on emails, documents, or books
  - Plagiarism and Phishing
- Application to short text (eg tweets) is a well-known problem
  - Beyond scope of project
  - Basic proof of concept

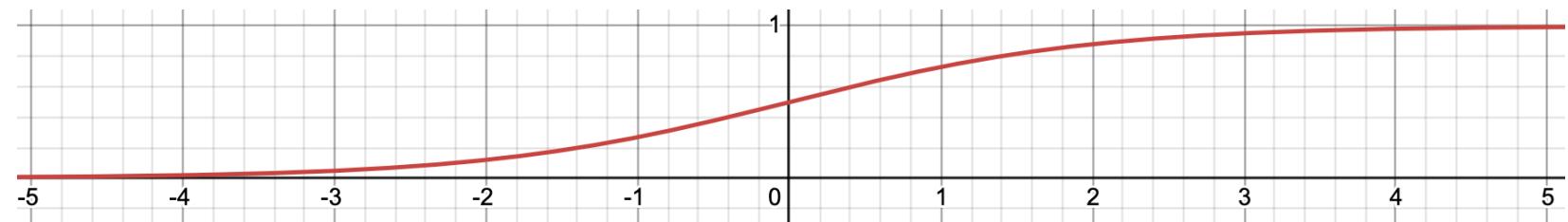
# Adaptation - Topic

- Messages tagged using HuggingFace Zero-Shot Classification model into knowledge areas
  - <RICE, SELLING>
  - Classify data into classes – unsupervised
- More questions asked about a topic, the more unknown
- Gets more accurate over time as full crop cycle repeated



# Adaptation - Topic

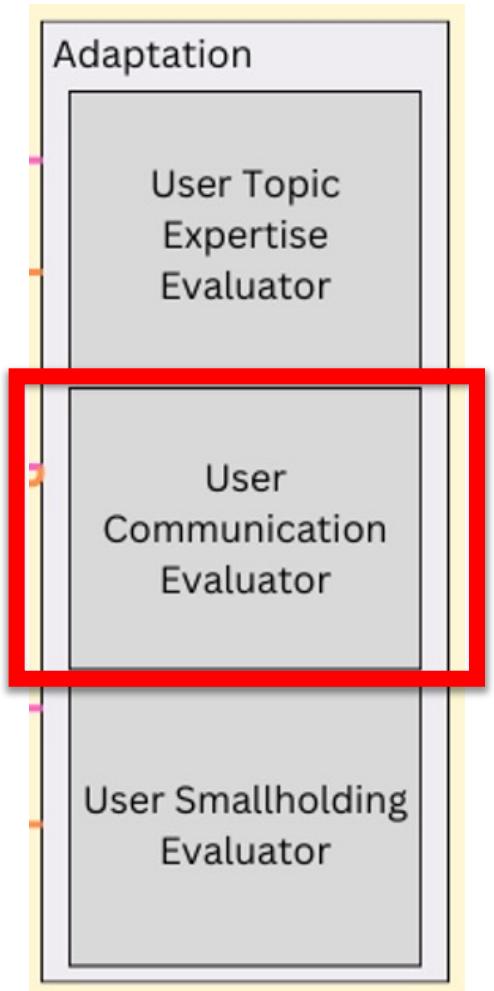
- Uses non-linear sigmoid function to provide numerical quantification of knowledge
- 0.5, neutral, 1, most knowledgable
- Notifications and summary



```
fun scaleProbability(actual: Double, bound: Double, inverseResult: Boolean = false): Double {  
    if (inverseResult) return 1.0 - (1.0 / (1.0 + exp(x: -0.1 * (actual - bound))))  
    return 1.0 / (1.0 + exp(x: -1 * (actual - bound)))  
}
```

# Adaptation - User

- Use pre-trained HuggingFace Models for Age and Gender.
- Literacy is scale of 0.0 to 100.0 (best)



# Adaptation - User

Weighted  
through  
test

Use JLangTool to calculate  
grammar and spelling mistakes

```
// Calculate average word count per message
val averageWordCount = messages?.map { it.message.length }?.let { calculateWordCountScore(it.average()) }

// Calculate average number of errors per message - weighted
val averageErrorsPerMessage = (messages?.map { errorsInMessage(it.message) }?.average() ?: 0.0) * 6

// Calculate average message readability - weighted
val averageReadability = (messages?.map { messageReadability(it.message) }?.average() ?: 0.0) * 2

// Calculate average vocabulary complexity (Type-Token Ratio)
val averageVocabularyComplexity = messages?.map { calculateTypeTokenRatio(it.message) }?.average()

// Combine individual metrics to calculate overall user literacy level
val literacy = ((averageWordCount ?: 0.0) +
    (averageErrorsPerMessage) +
    (averageReadability) +
    (averageVocabularyComplexity ?: 0.0) / 10.0).toFloat()
```

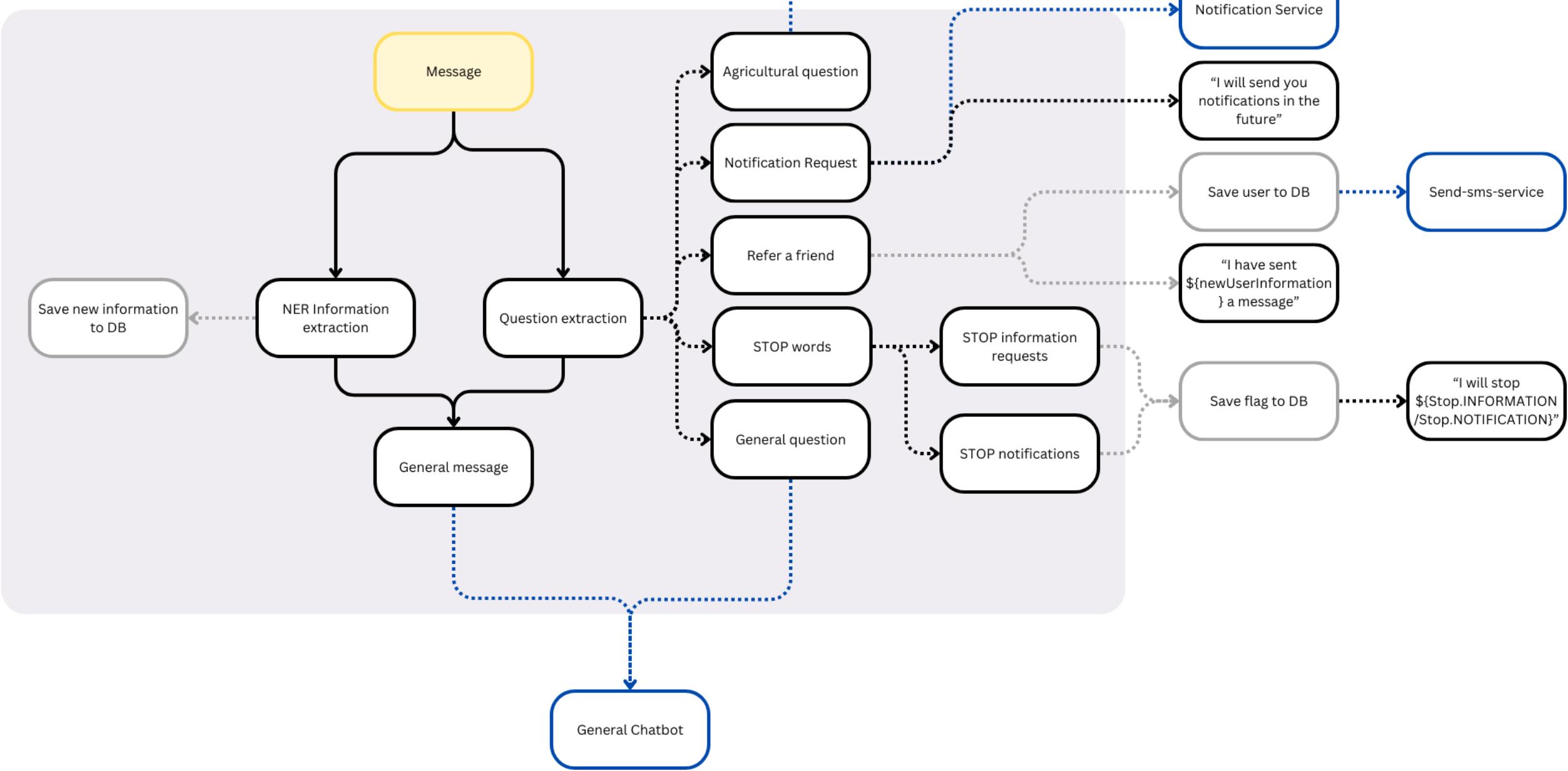
Type-Token Ratio to obtain  
vocabulary complexity and variability

Use Flesch Reading-ease  
Test to calculate readability

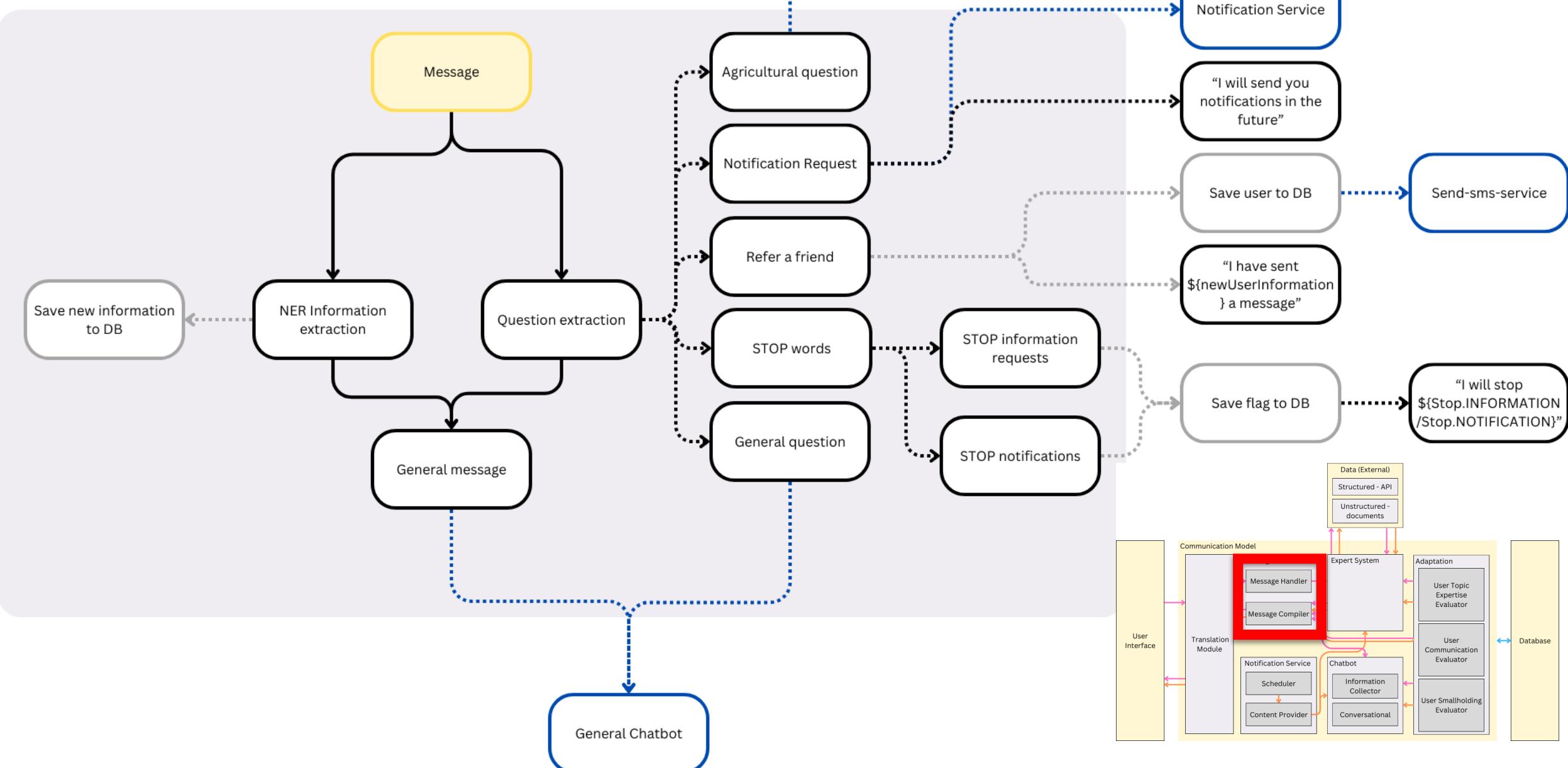
# Adaptation -User

- Use OpenAI
- Could not find studies on improving uptake of suggestions in messaging based on short-text authorship characteristics
  - Much scope for future work
- However, even rudimentary approach seems to have had an effect

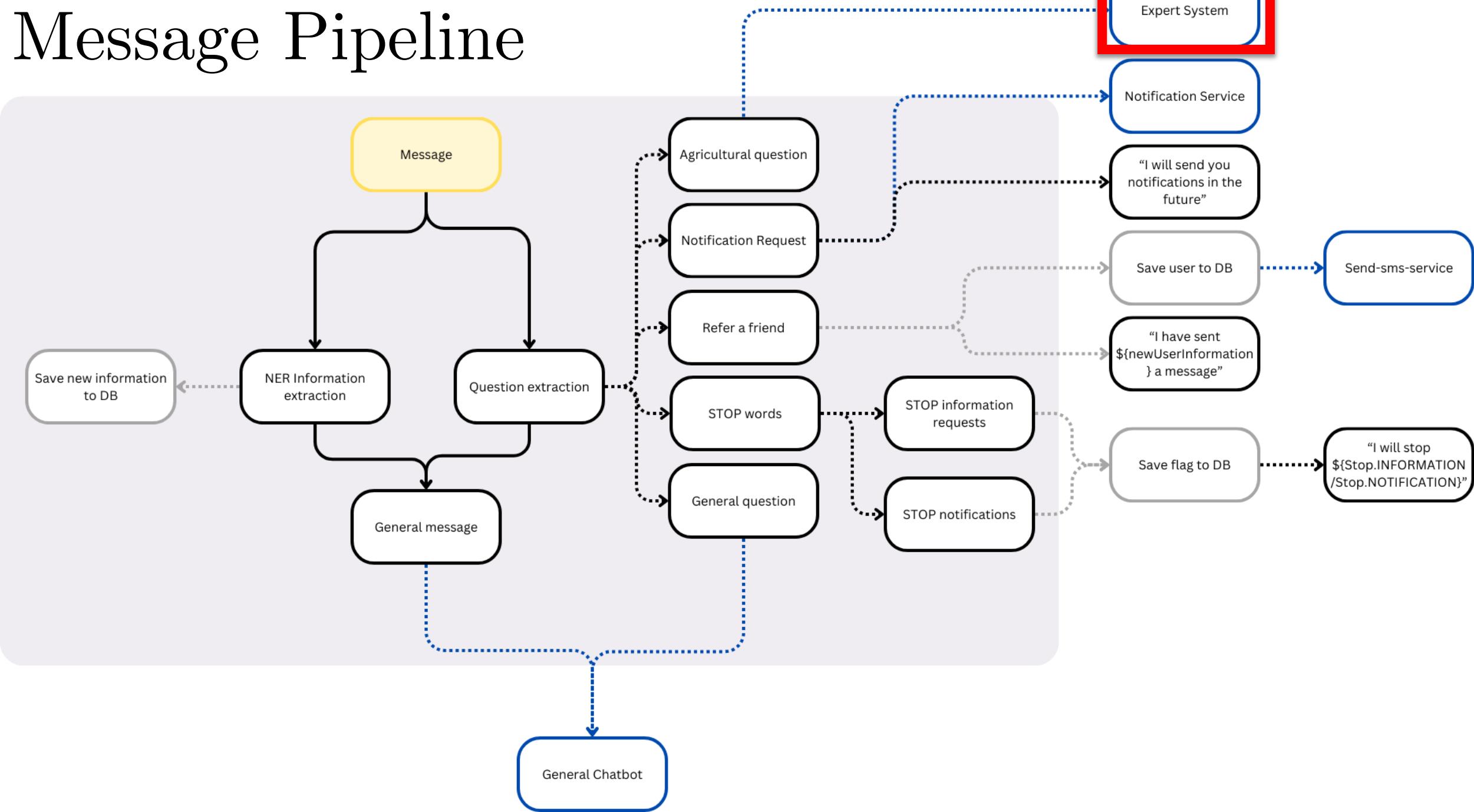
# Message Pipeline



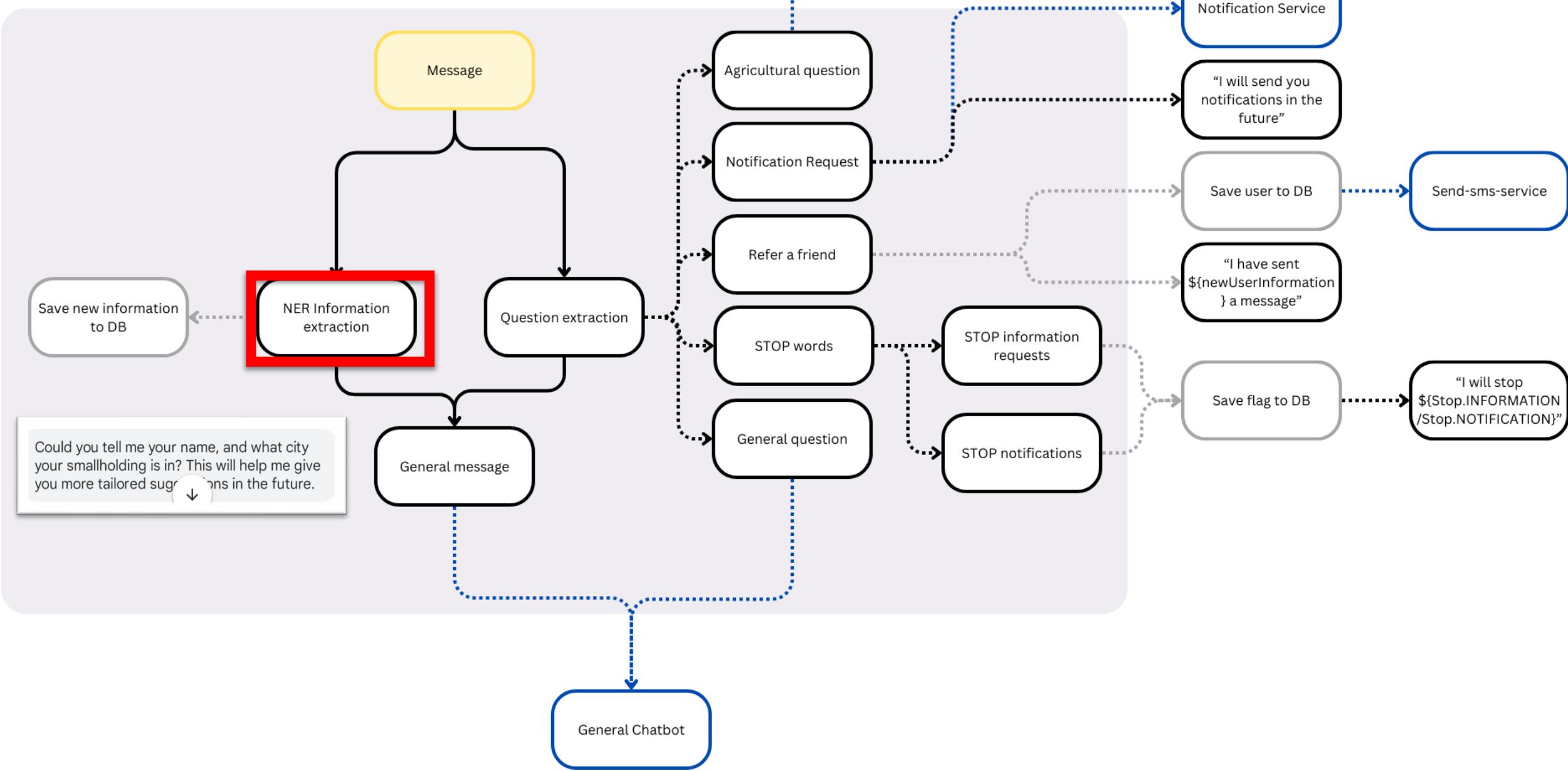
# Message Pipeline



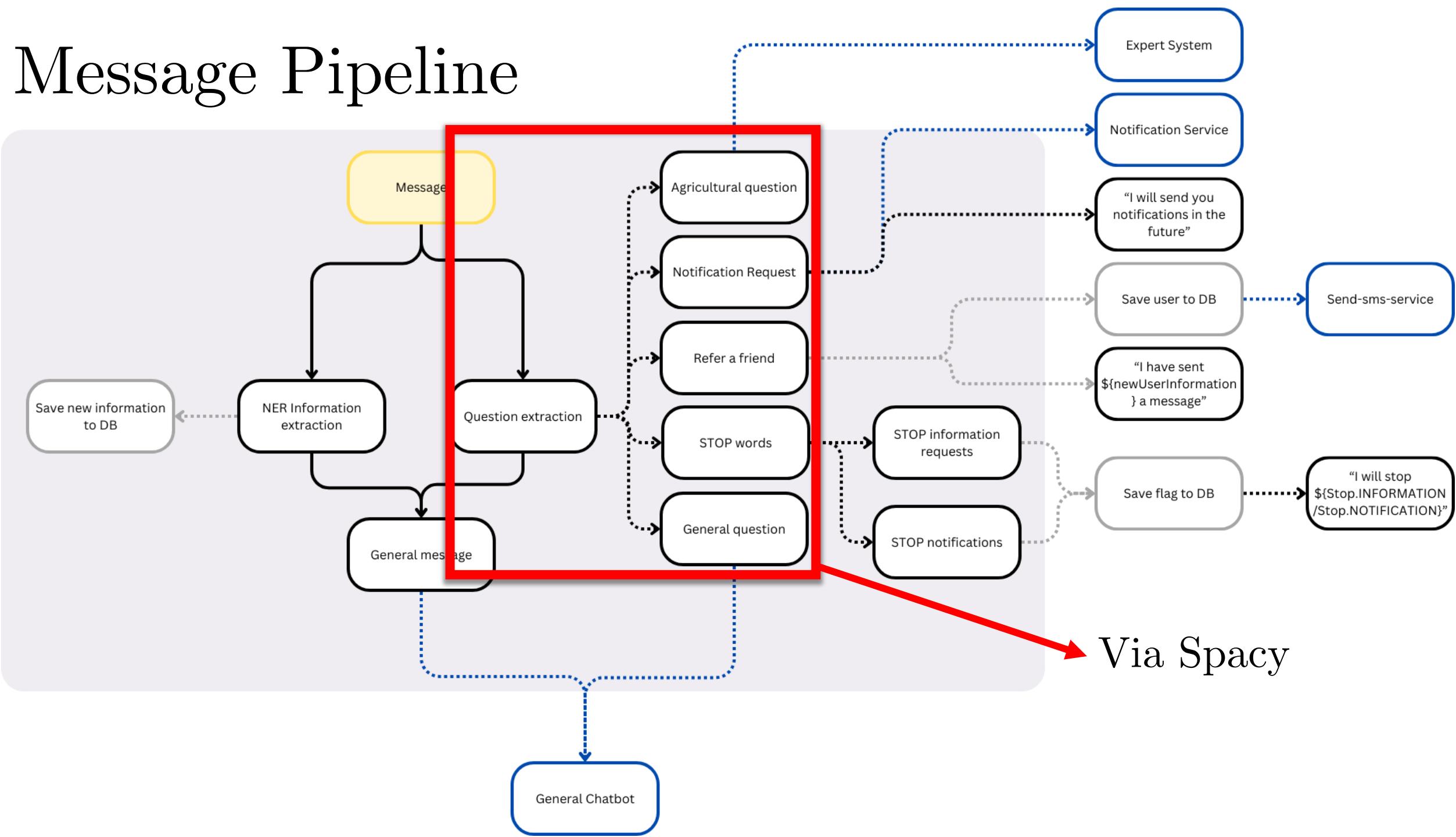
# Message Pipeline



# Message Pipeline



# Message Pipeline



# Message Pipeline

- Explicit and implicit questions
- Doesn't have to follow punctuation rules

```
nlp = spacy.load("en_core_web_sm")

# usage: new *

def isQuestion(sentence):
    """
    Determine if a sentence is a direct question or implies a question based on syntactic structure.
    """

    # Check for sentences starting with wh-words
    if sentence[0].text.lower() in ["how", "what", "when", "where", "which", "who", "whom", "whose", "why"]:
        return True

    for token in sentence:
        # Check for inversion: verb + subject
        if token.dep_ == "ROOT" and token.tag_ == "VB":
            for child in token.children:
                if child.dep_ == "nsubj":
                    return True
        # Check for implied questions
        if token.dep_ == "xcomp" and token.head.text.lower() in ["want", "wonder", "ask", "know"]:
            return True
    return False

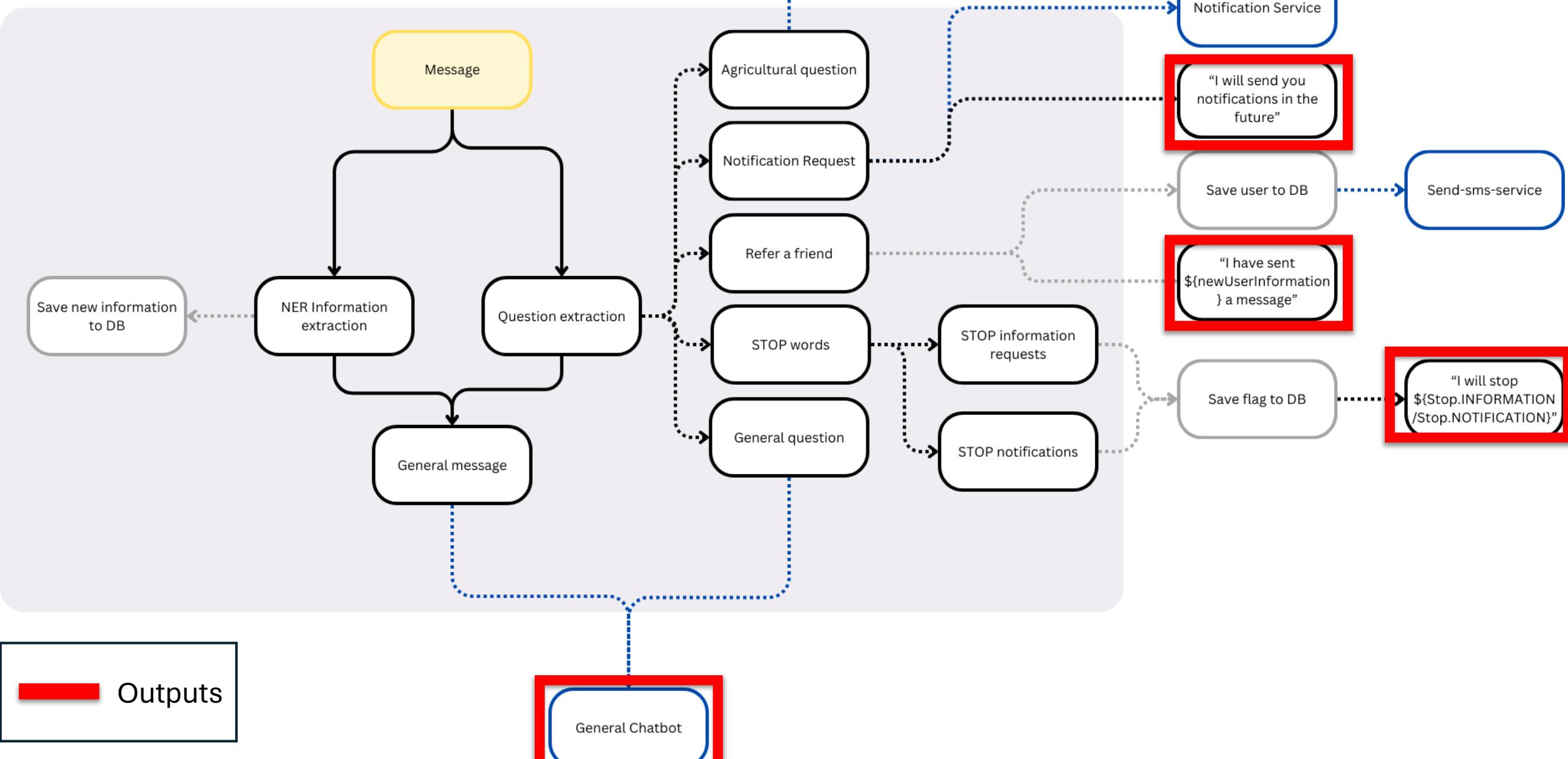
text = "This is an example text. What is your name? Where are you from? I want to know when to plant seeds."
doc = nlp(text)

questions = []

for sent in doc.sents:
    if isQuestion(sent):
        questions.append(sent.text)

print(questions)
```

# Extracting Data



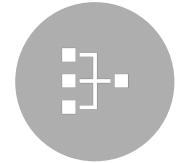
# Contents



Motivations



Research



System



Expert  
System



Adaptiveness



Other  
Achievements



Conclusions

# Kotlin-Python Interop - annotation

- Custom spring annotation for ease of use

```
▲ Jacqueline Dobreva-Skevington
@PythonFunction(functionName: "collect", scriptName: "InformationCollector.py")
fun collect(userDetails: List<UserDetails>): String? {
    return execute(::collect, userDetails)
}
```

- Just need to include file and function name

# Kotlin-Python Interop - communication

```
inline fun <reified T> executeProgram(scriptName: String, functionName: String, arguments: List<Any?>): T {
    val programRunUID = UUID.randomUUID().toString()
    writeArgumentsToFile(arguments, filename: "$programRunUID.args.json")
    val outputFile = File(getPathForFile(filename: "$programRunUID.out.txt"))
    outputFile.writeText(text: "")
    val process = ProcessBuilder()
        .command(
            mutableListOf(
                "python3",
                getPythonProgram(scriptName),
                programRunUID,
                functionName,
                getPathForFile(filename: ""))
        )
    .redirectOutput(outputFile)
    .redirectError(outputFile)
    .start()
    val job = CoroutineScope(Dispatchers.IO).async { this: CoroutineScope
        var currentLines = 0
        while (true) {
            delay(timeMillis: 100)
            val lines = outputFile.readLines()
            for (i in currentLines .. until < lines.size) {
                println(lines[i])
            }
            currentLines = lines.size
        }
    }
    val result = process.onExit().join()
    runBlocking { this: CoroutineScope
        job.cancelAndJoin()
    }
    return readResultFromFile<T>(filename: "$programRunUID.result.json")
}
```

- Uses JSON to communicate results and arguments between Kotlin and Python programs
- Asynchronous polling for real-time output from the Python process

# Kotlin-Python Interop - errors

- Handles errors from Python process in Kotlin system to allow for debugging, as well as targeted error handling
- Will be published a standalone library

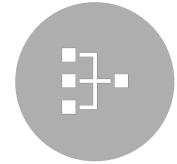
# Contents



Motivations



Research



System



Expert  
System



Adaptiveness



Other  
Achievements



Conclusions

# Statistics

- Takes <8 seconds to formulate response
  - Comparable
- Guesses age, gender, literacy with ~80% accuracy
- 1.3 times as likely to implement user characteristic tailored suggestions

# Successes

- Designed an expandable system that supports multiple crops, for all stages of the crop cycle, globally
- Extends existing work on using RAG in the Agricultural context (tailoring)
- Proposes new suggestibility concept
- Multiple libraries to be published
- Own successes on the allotment

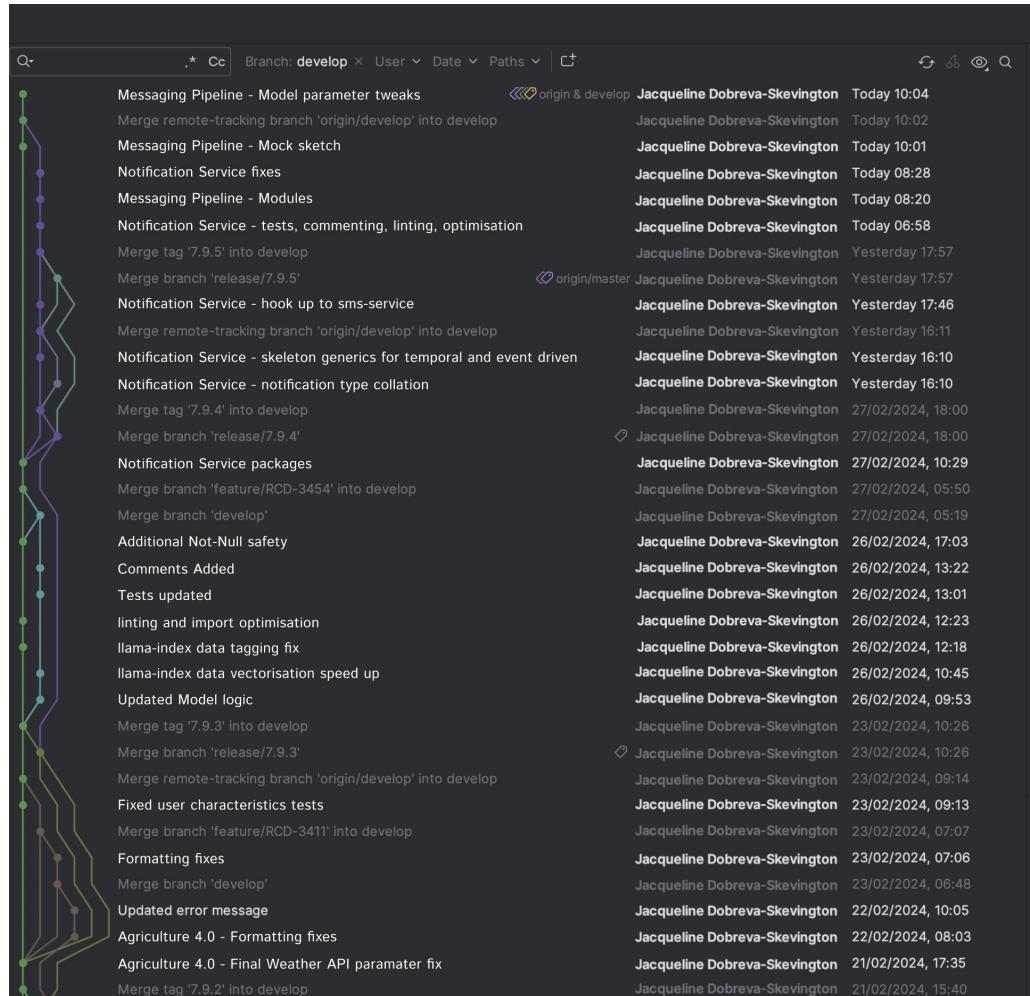
# Future work

- Local dialect translation
- IVR option
- Specific work on short-text authorship characteristics
- Evaluation of culture rules and social biases to influence responses – more documents

# Project Management



# Project Management



# Project Management



Branch: develop		
Q	*	Cc
Messaging Pipeline - Model parameter tweaks	origin & develop	Jacqueline Dobreva-Skevington Today 10:04
Merge remote-tracking branch 'origin/develop' into develop		Jacqueline Dobreva-Skevington Today 10:02
Messaging Pipeline - Mock sketch		Jacqueline Dobreva-Skevington Today 10:01
Notification Service fixes		Jacqueline Dobreva-Skevington Today 08:28
Messaging Pipeline - Modules		Jacqueline Dobreva-Skevington Today 08:20
Notification Service - tests, commenting, linting, optimisation		Jacqueline Dobreva-Skevington Today 06:58
Merge tag '7.9.5' into develop		Jacqueline Dobreva-Skevington Yesterday 17:57
Merge branch 'release/7.9.5'	origin/master	Jacqueline Dobreva-Skevington Yesterday 17:57
Notification Service - hook up to sms-service		Jacqueline Dobreva-Skevington Yesterday 17:46
Merge remote-tracking branch 'origin/develop' into develop		Jacqueline Dobreva-Skevington Yesterday 16:11
Notification Service - skeleton generics for temporal and event driven		Jacqueline Dobreva-Skevington Yesterday 16:10
Notification Service - notification type collation		Jacqueline Dobreva-Skevington Yesterday 16:10
Merge tag '7.9.4' into develop		Jacqueline Dobreva-Skevington 27/02/2024, 18:00
Merge branch 'release/7.9.4'		Jacqueline Dobreva-Skevington 27/02/2024, 18:00
Notification Service packages		Jacqueline Dobreva-Skevington 27/02/2024, 10:29
Merge branch 'feature/RCD-3454' into develop		Jacqueline Dobreva-Skevington 27/02/2024, 05:50
Merge branch 'develop'		Jacqueline Dobreva-Skevington 27/02/2024, 05:19
Additional Not-Null safety		Jacqueline Dobreva-Skevington 26/02/2024, 17:03
Comments Added		Jacqueline Dobreva-Skevington 26/02/2024, 13:22
Tests updated		Jacqueline Dobreva-Skevington 26/02/2024, 13:01
linting and import optimisation		Jacqueline Dobreva-Skevington 26/02/2024, 12:23
llama-index data tagging fix		Jacqueline Dobreva-Skevington 26/02/2024, 12:18
llama-index data vectorisation speed up		Jacqueline Dobreva-Skevington 26/02/2024, 10:45
Updated Model logic		Jacqueline Dobreva-Skevington 26/02/2024, 09:53
Merge tag '7.9.3' into develop		Jacqueline Dobreva-Skevington 23/02/2024, 10:26
Merge branch 'release/7.9.3'		Jacqueline Dobreva-Skevington 23/02/2024, 10:26
Merge remote-tracking branch 'origin/develop' into develop		Jacqueline Dobreva-Skevington 23/02/2024, 09:14
Fixed user characteristics tests		Jacqueline Dobreva-Skevington 23/02/2024, 09:13
Merge branch 'feature/RCD-3411' into develop		Jacqueline Dobreva-Skevington 23/02/2024, 07:07
Formatting fixes		Jacqueline Dobreva-Skevington 23/02/2024, 07:06
Merge branch 'develop'		Jacqueline Dobreva-Skevington 23/02/2024, 06:48
Updated error message		Jacqueline Dobreva-Skevington 22/02/2024, 10:05
Agriculture 4.0 - Formatting fixes		Jacqueline Dobreva-Skevington 22/02/2024, 08:03
Agriculture 4.0 - Final Weather API parameter fix		Jacqueline Dobreva-Skevington 21/02/2024, 17:35
Merge tag '7.9.2' into develop		Jacqueline Dobreva-Skevington 21/02/2024, 15:40

Laptop broke

# Text ‘S’ for Smallholding

A Personalised Unified SMS-based Agricultural Extension Service  
for Small-Scale Farmers