CAPSTONE PROJECT

Graciela Diwa


I SAW A DINOSAUR!
I SAW A DRAGON!
I SAW A SCARY FISH IN THE RIVER!

Some things can be hard to believe.

They can be even harder to imagine.

CAPSTONE PROJECT

APPLYING NLP TECHNIQUES TO EXTRACT KEY DESCRIPTORS FOR AI IMAGE GENERATION

Graciela Diwa

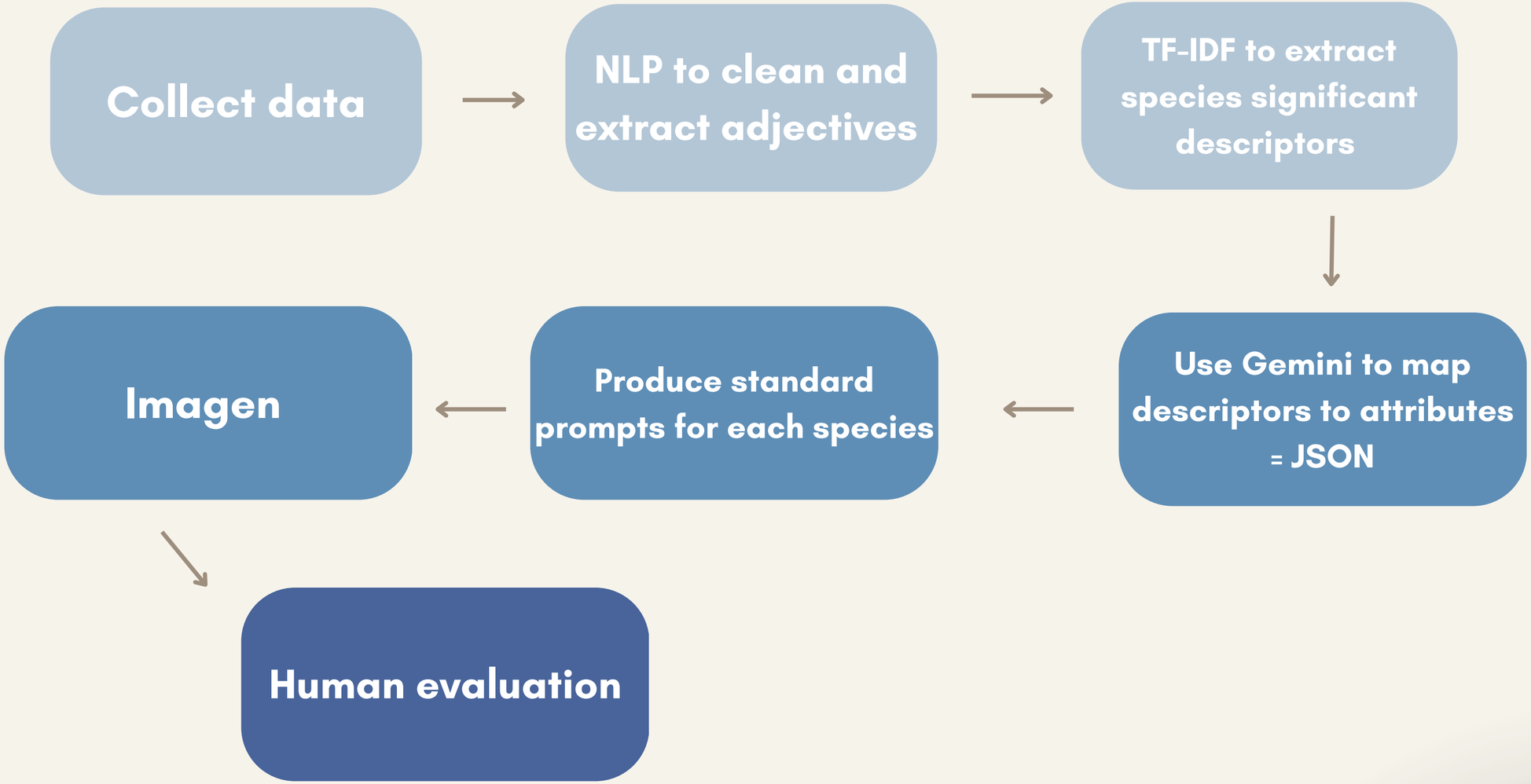
Sensational vs Scientific

Sensational vs Scientific

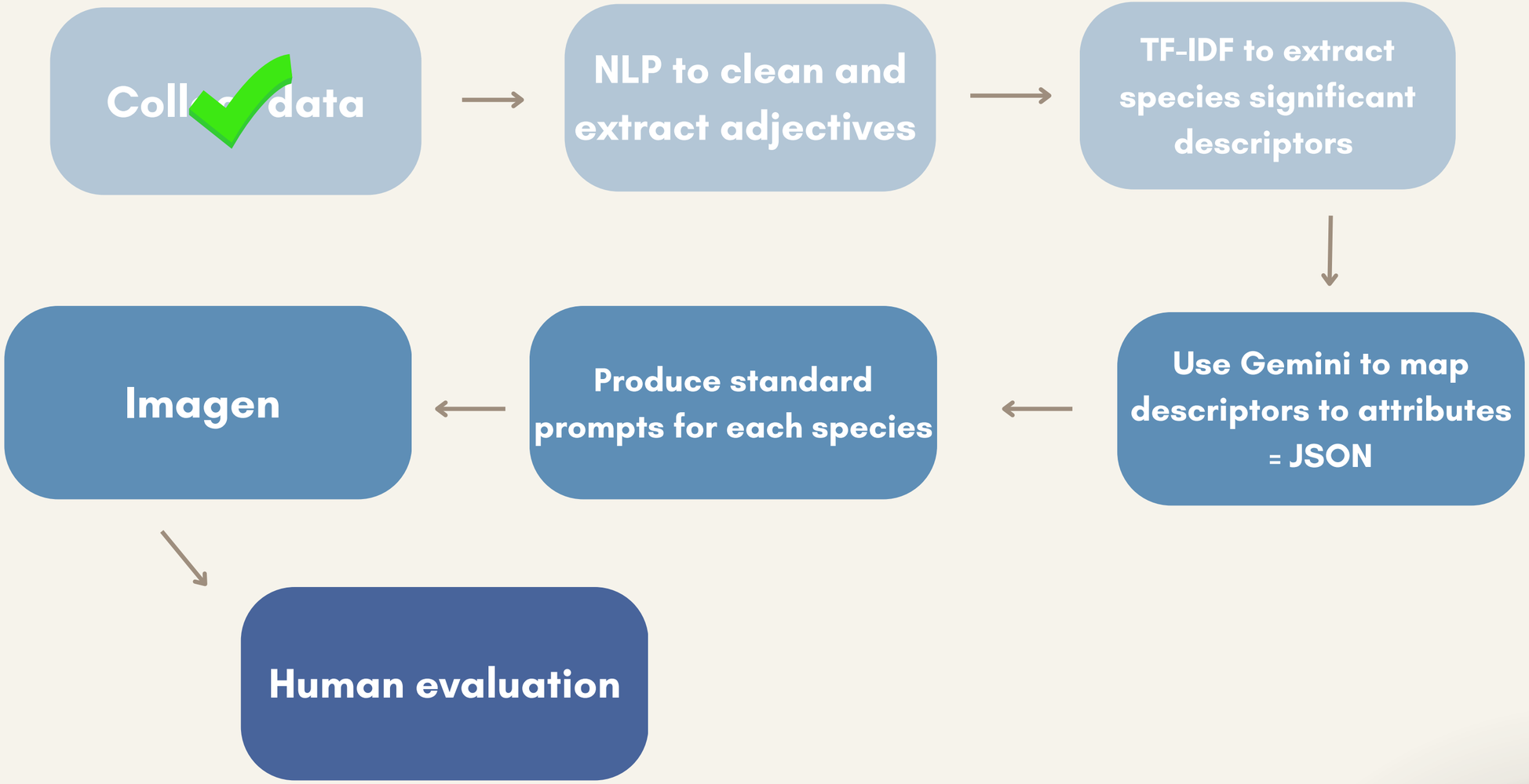
Sensational vs Scientific

FishBase

PROJECT STAGES



PROJECT STAGES



NLP to extract descriptors

- Normalise the data
  - lowercase, removing punctuation, timestamps
  - a lot of RegEx
- Reduce noise
  - embedded semantic filtering*
  - removing stopwords
  - POS or adj list
- Clean, clean, clean
- ..and clean some more

chunk the text, dynamic → chunks

create variants and concept of the thing → queries

cosine similarity to keep only relevant chunks → threshold, keywords

NLP to extract descriptors

- Normalise the data
  - lowercase, removing punctuation, timestamps
  - a lot of RegEx
- Reduce noise
  - embedded semantic filtering
  - removing stopwords
  - POS or adj list *
- Clean, clean, clean
- ..and clean some more

Using spaCy POS tagging ‘ADJ’ to keep descriptive words only

adjectives list to filter out non descriptive words

NLP to extract descriptors

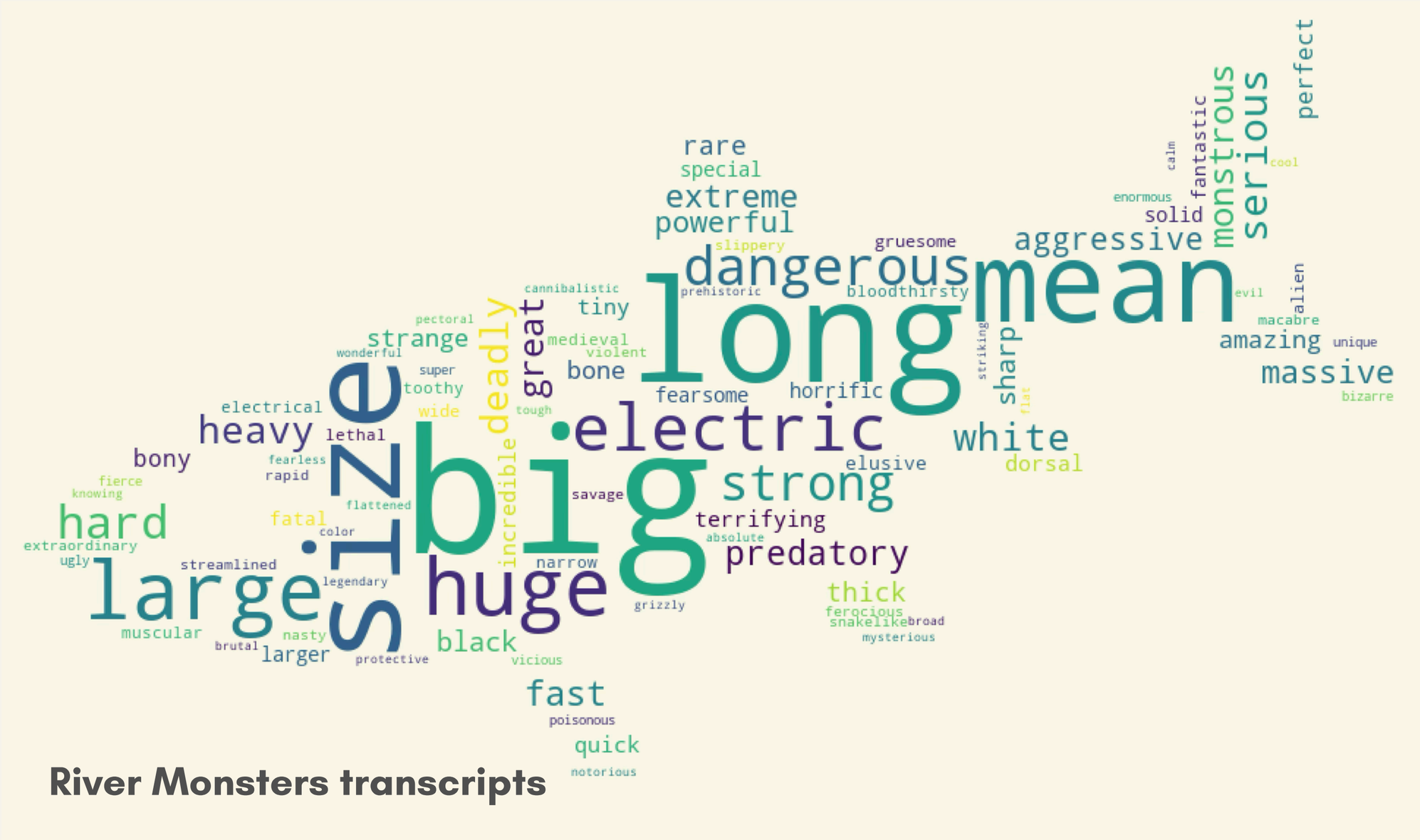
- Normalise the data
  - lowercase, removing punctuation, timestamps
  - a lot of RegEx
- Reduce noise
  - embedded semantic filtering
  - removing stopwords
  - POS or adj list *
- Clean, clean, clean
- ..and clean some more

Using spaCy POS tagging ‘ADJ’ to keep descriptive words only

adjectives list to filter out non descriptive words

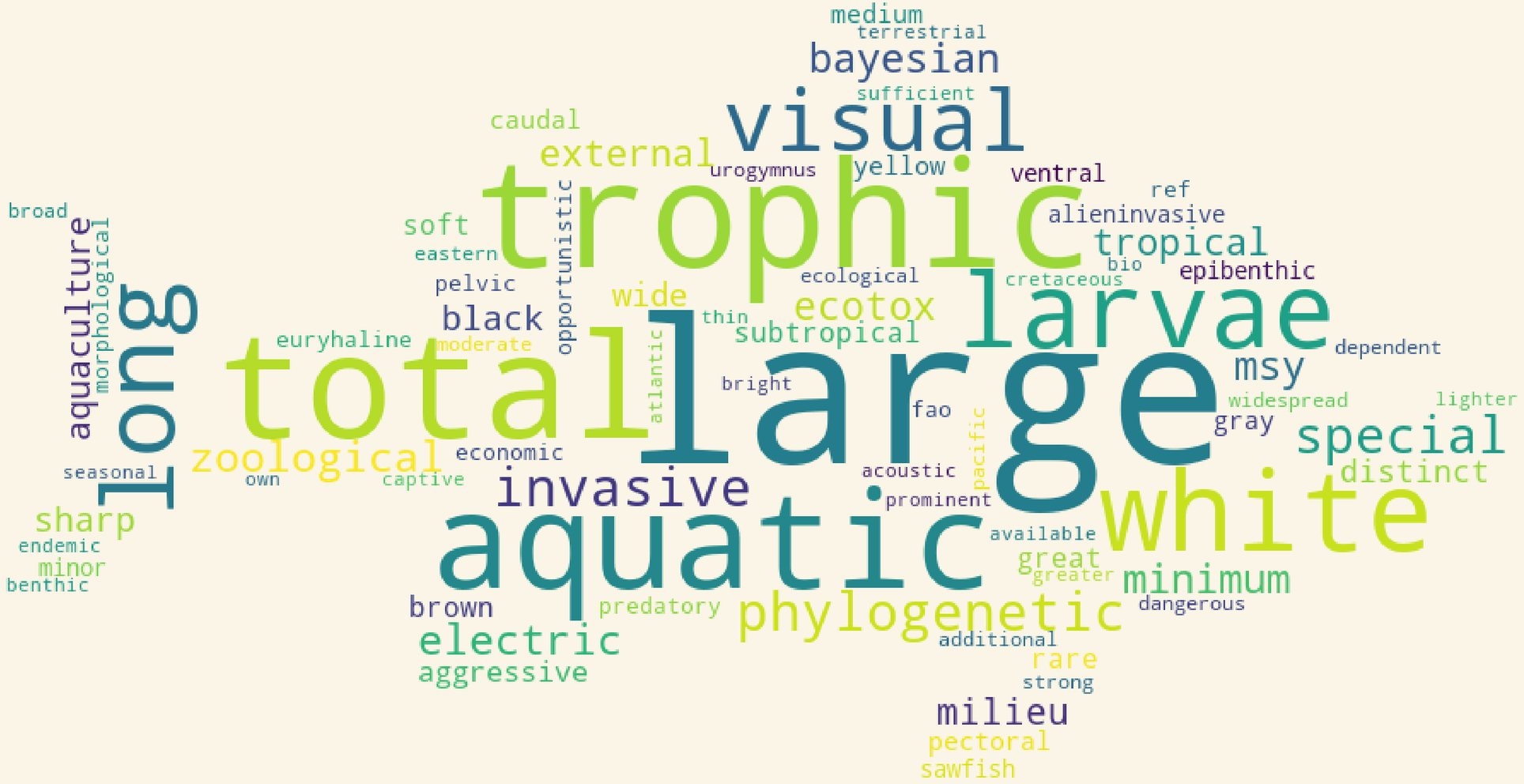
reduced more noise and specifically tagged words

sensational



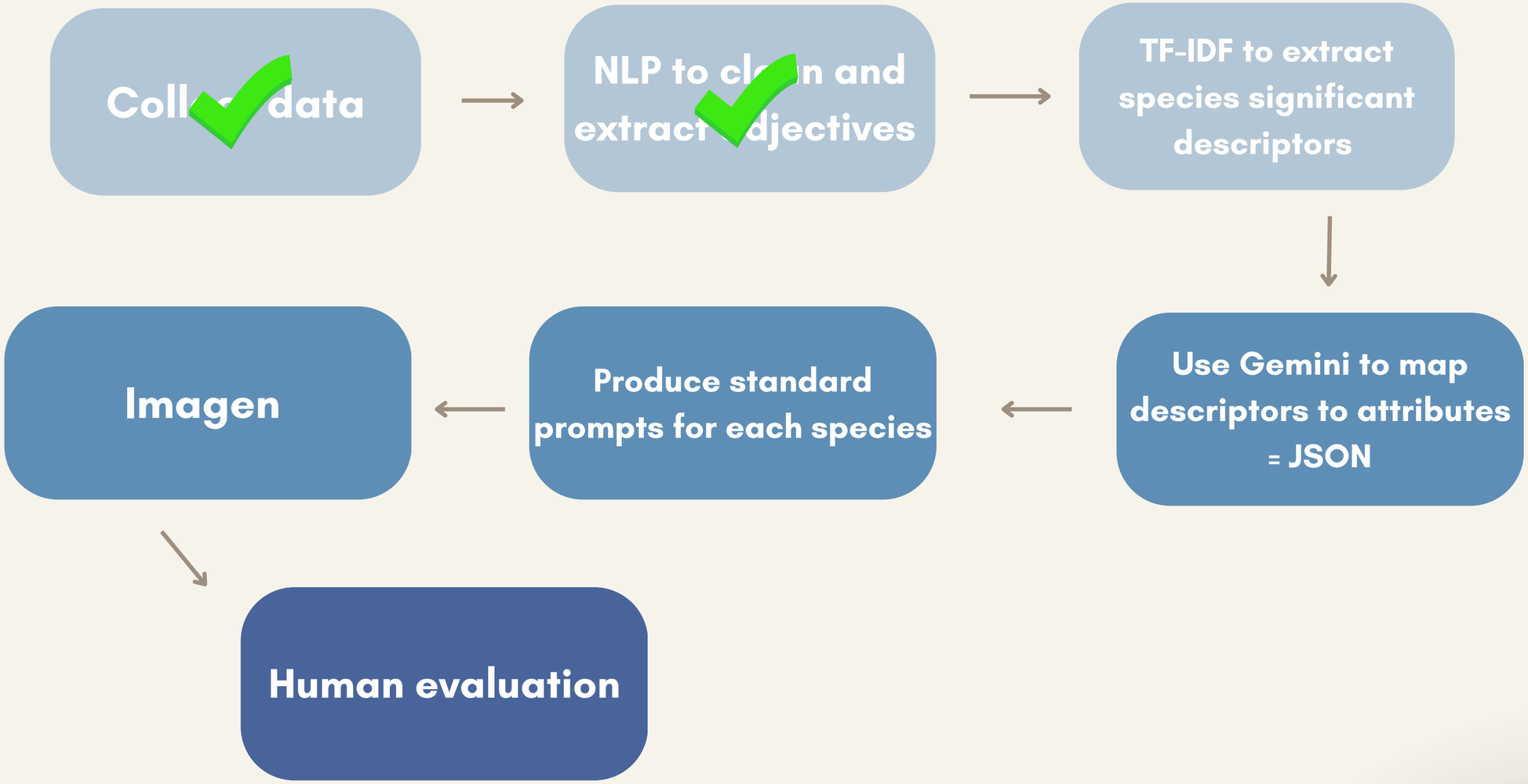
River Monsters transcripts

scientific

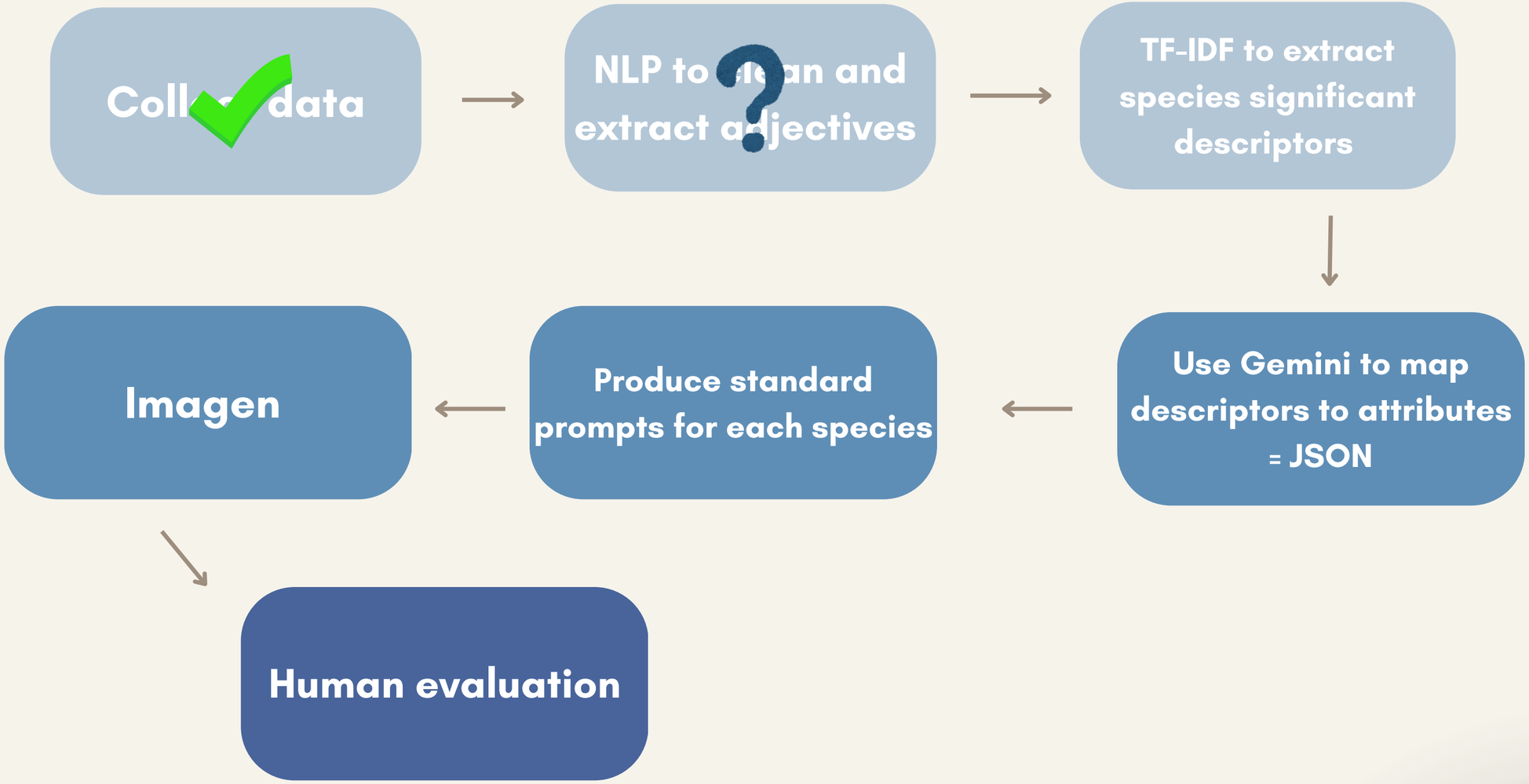


Wikipedia and Fishbase

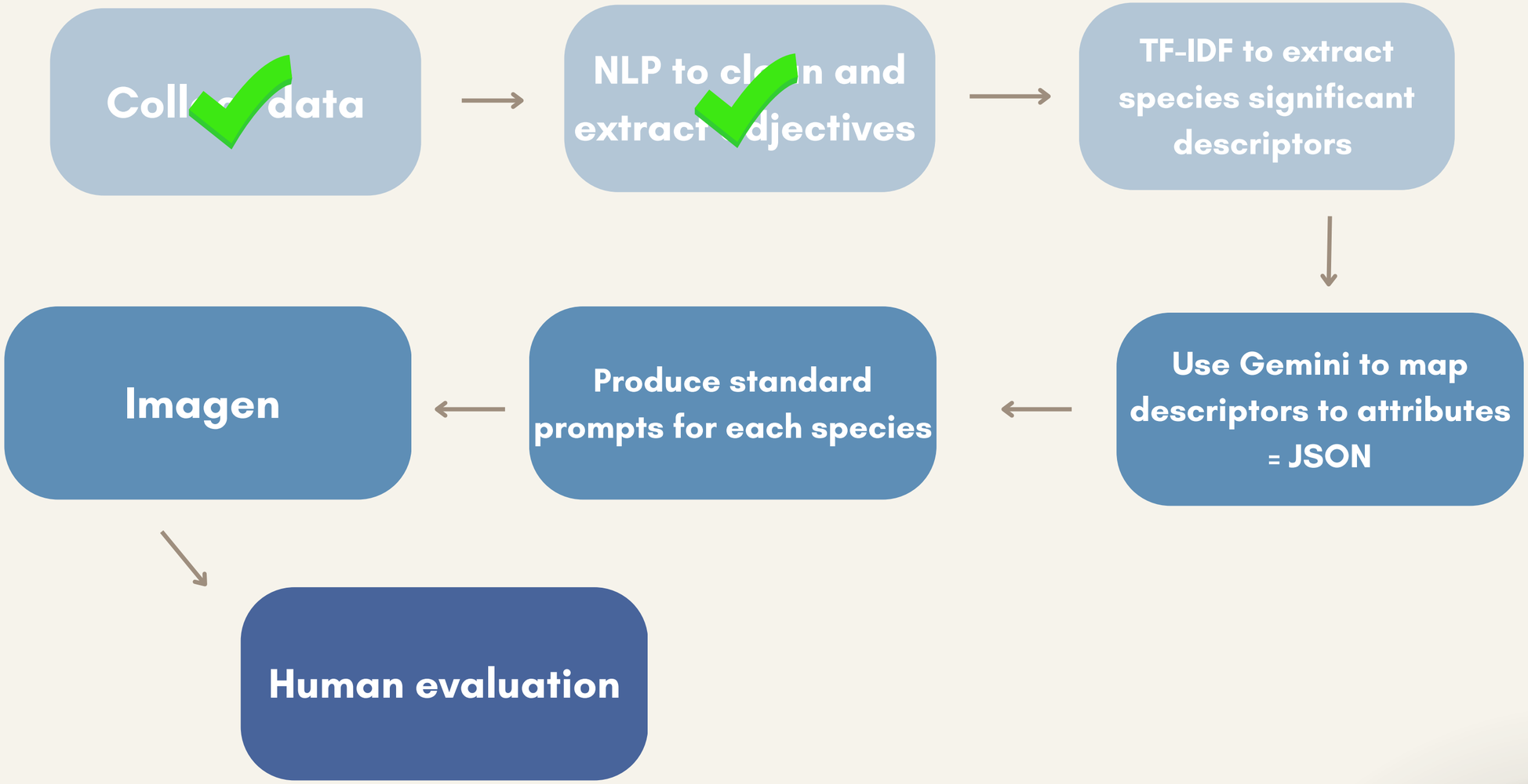
PROJECT STAGES



PROJECT STAGES



PROJECT STAGES



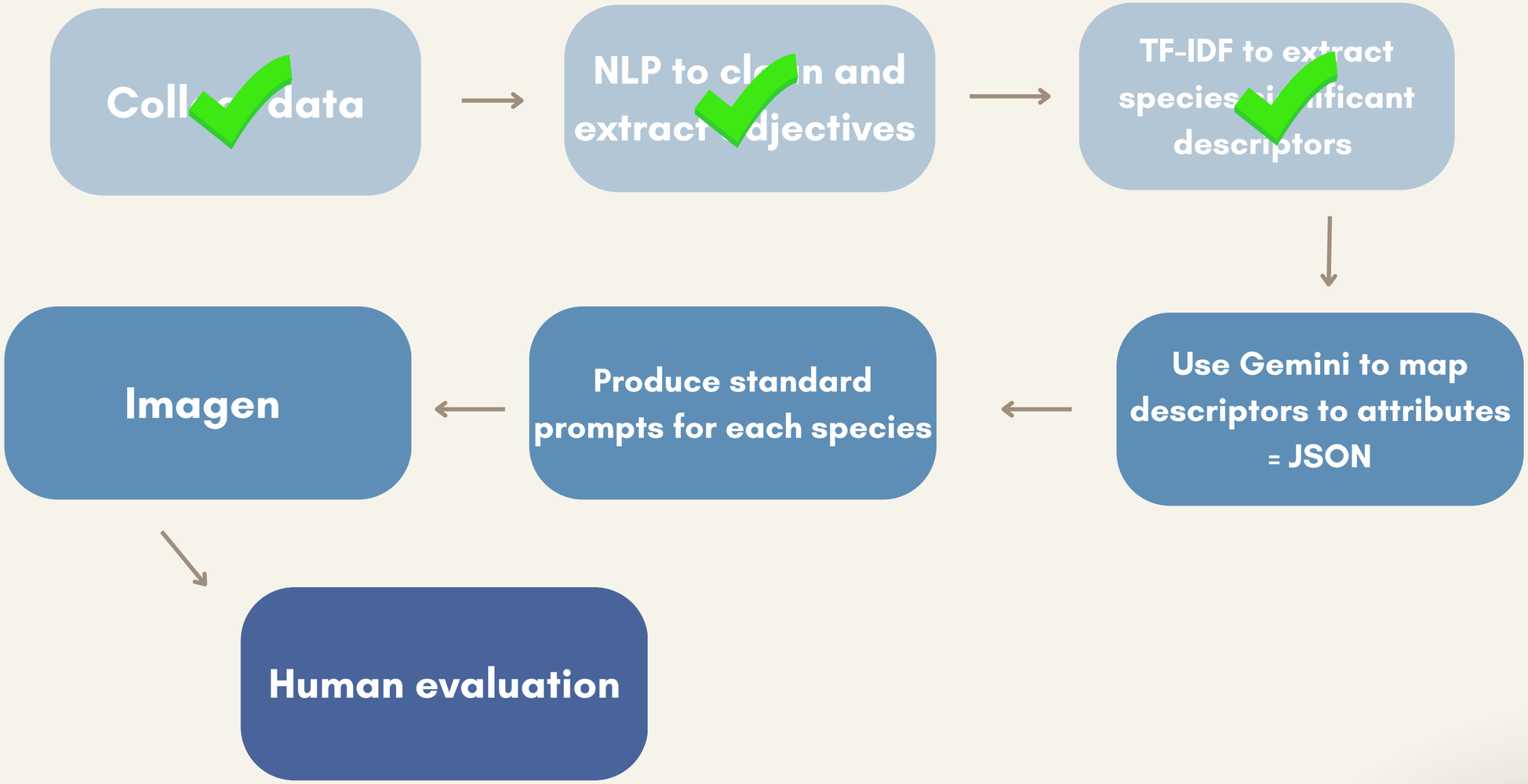
TF-IDF to extract species significant descriptors

- Keep descriptors that have high TF-IDF
  - more species specific
- Filter out common / generic descriptors

large, silvery, predatory

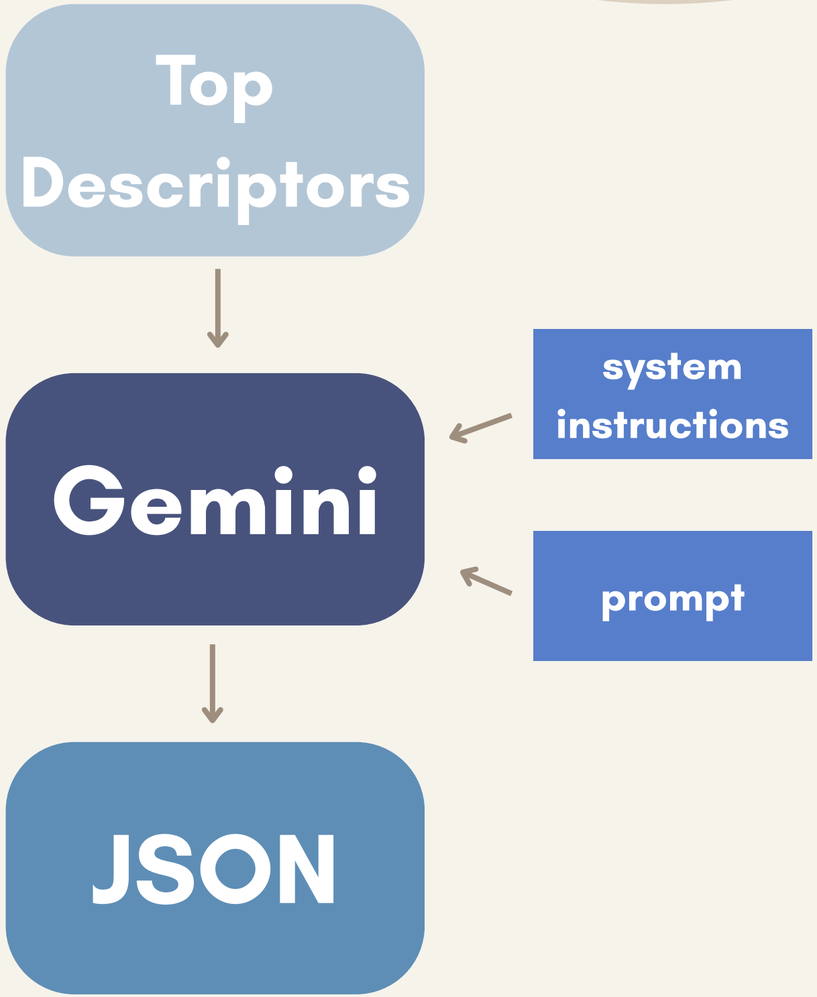
electric, nocturnal, muscular, spiny

PROJECT STAGES



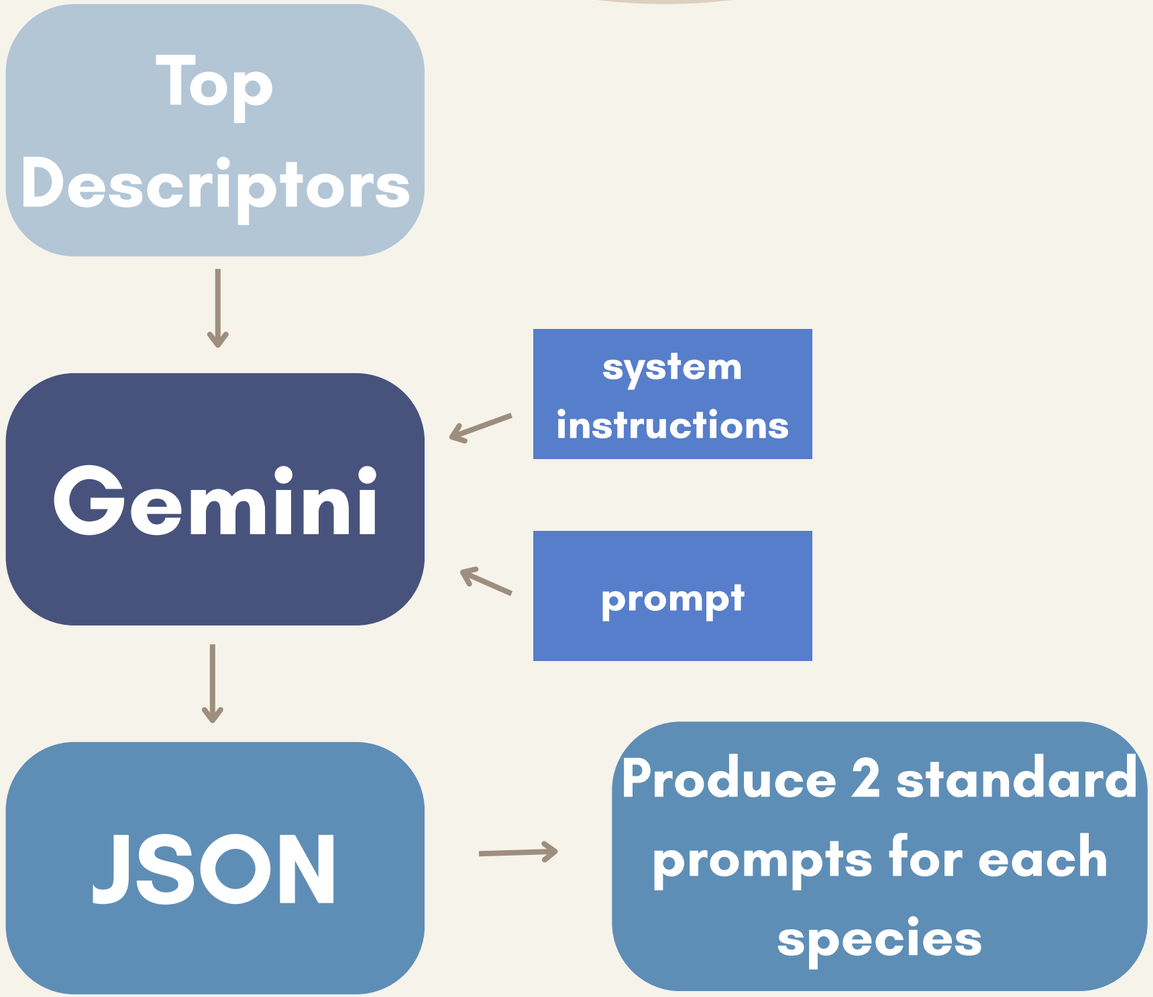
Use Gemini to map descriptors to attributes = JSON

- **Integrated LLM (Gemini) to map the top descriptors into JSON format**
- **eg.**
  - **‘sharp’** → ‘teeth’
  - **‘muscular’** → ‘body’
  - **no match** → category left blank

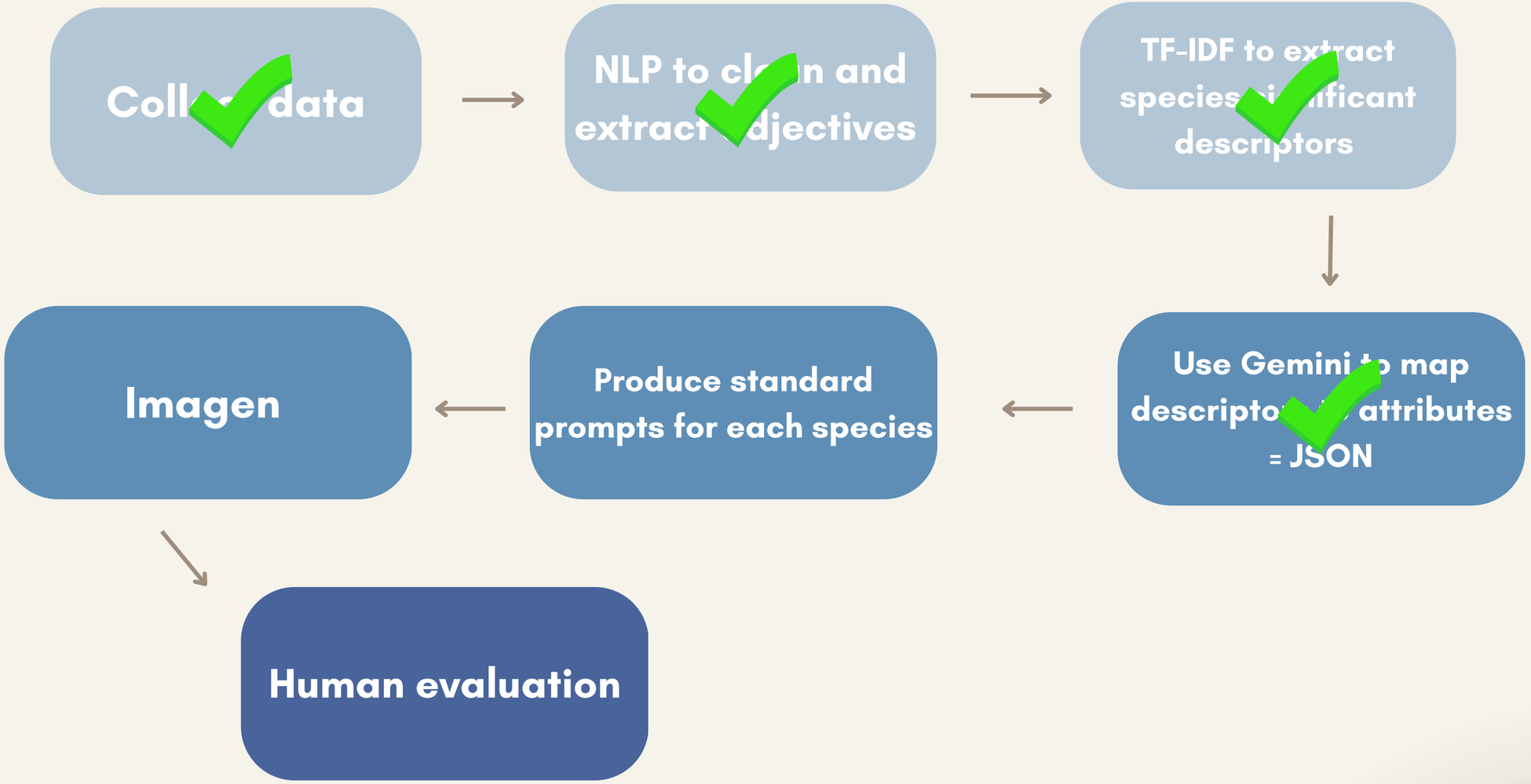


Use Gemini to map descriptors to attributes = JSON

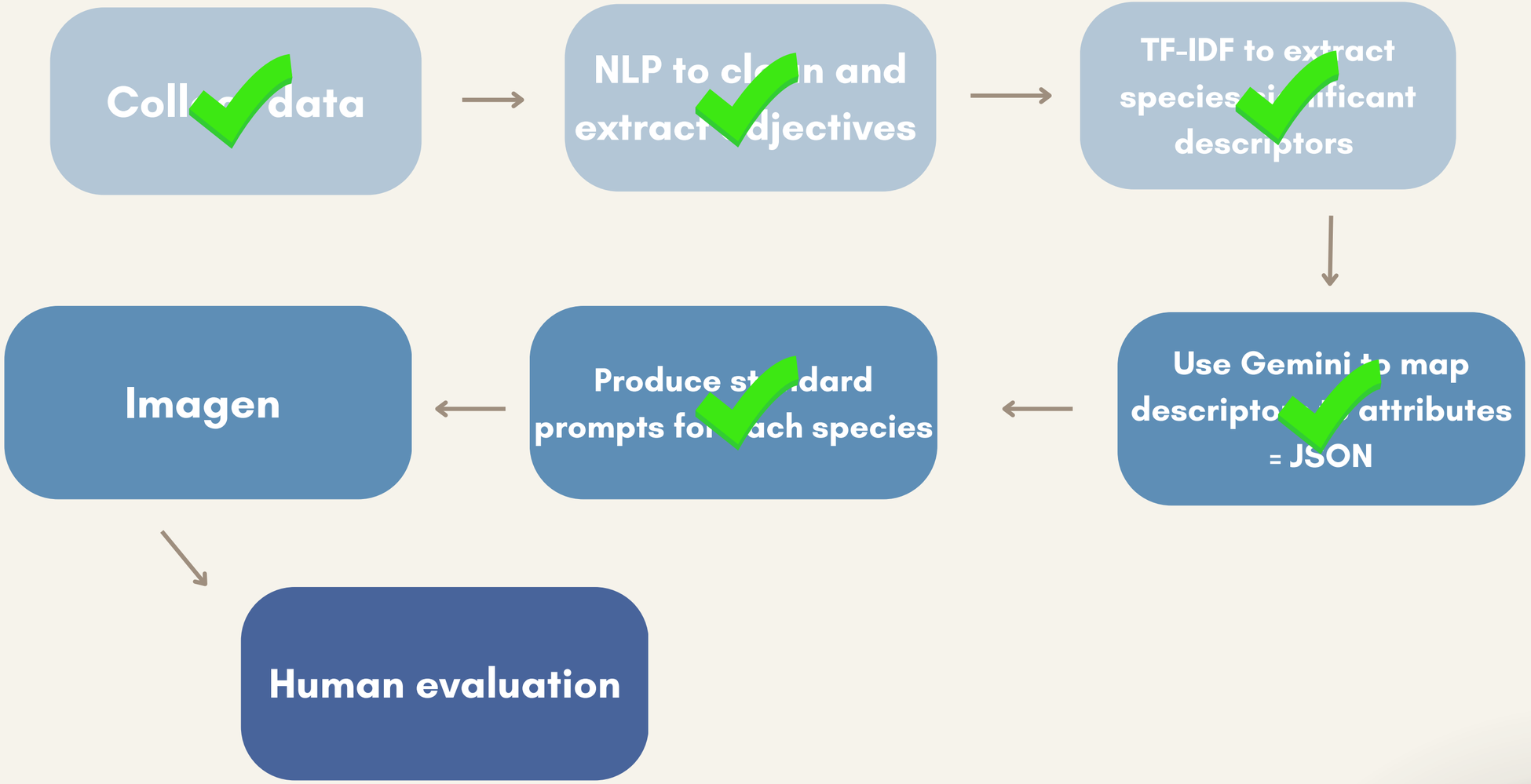
- **Integrated LLM (Gemini) to map the top descriptors into JSON format**
- **eg.**
  - **‘sharp’** → ‘teeth’
  - **‘muscular’** → ‘body’
  - **no match** → category left blank



PROJECT STAGES



PROJECT STAGES



Imagen

Feeding standardised prompts to Imagen

```
✅ Saved image as: goonch_catfish_transcript.png
goonch_catfish image generated!
Sleeping for 19.4 seconds...
✅ Saved image as: bagarius_yarrelli_wiki.png
bagarius_yarrelli image generated!
Sleeping for 18.0 seconds...
✅ Saved image as: alligator_gar_transcript.png
alligator_gar image generated!
Sleeping for 18.6 seconds...
✅ Saved image as: atractosteus_spatula_wiki.png
atractosteus_spatula image generated!
```

Imagen

Feeding standardised prompts to Imagen

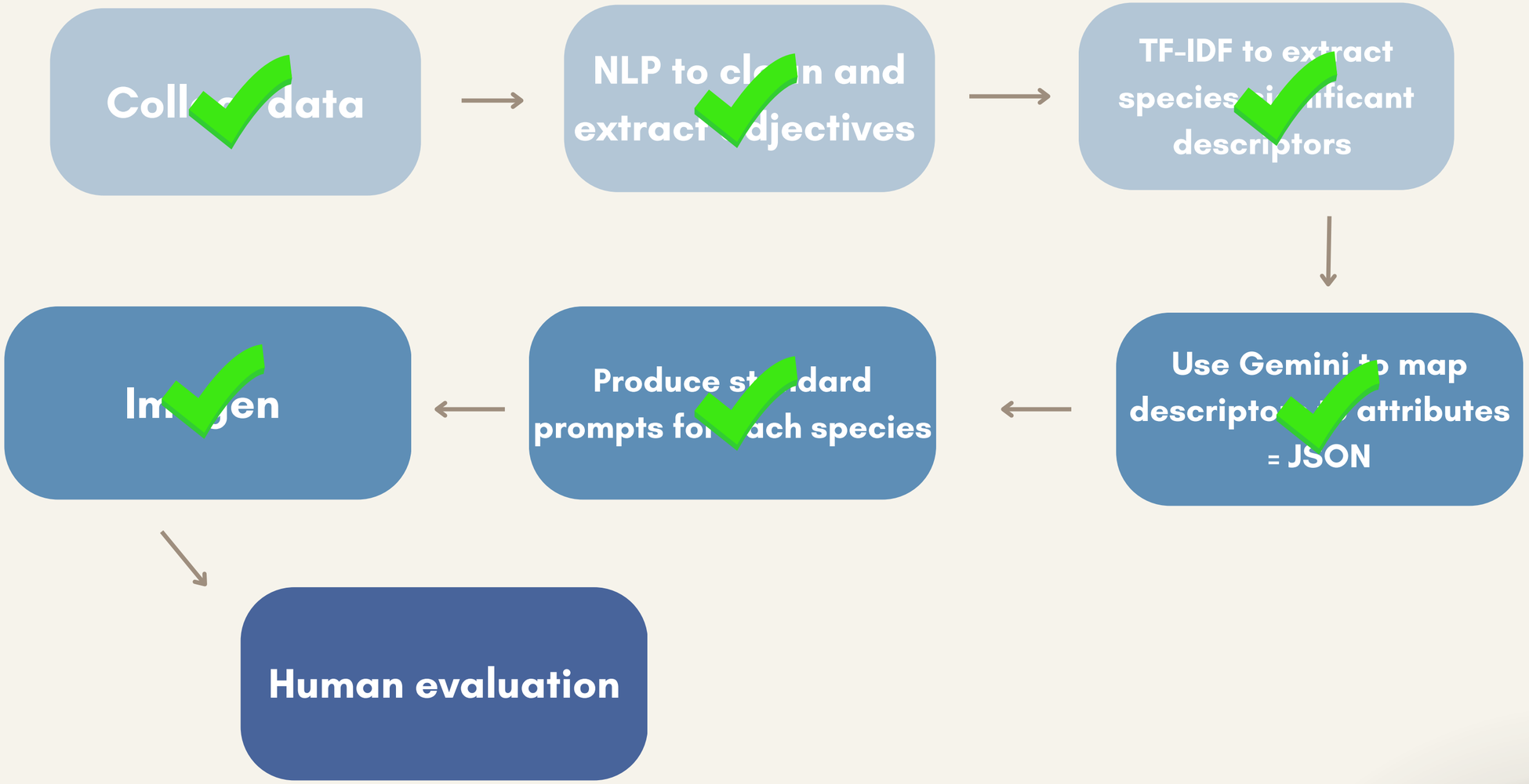
Imagen

Feeding standardised prompts to Imagen

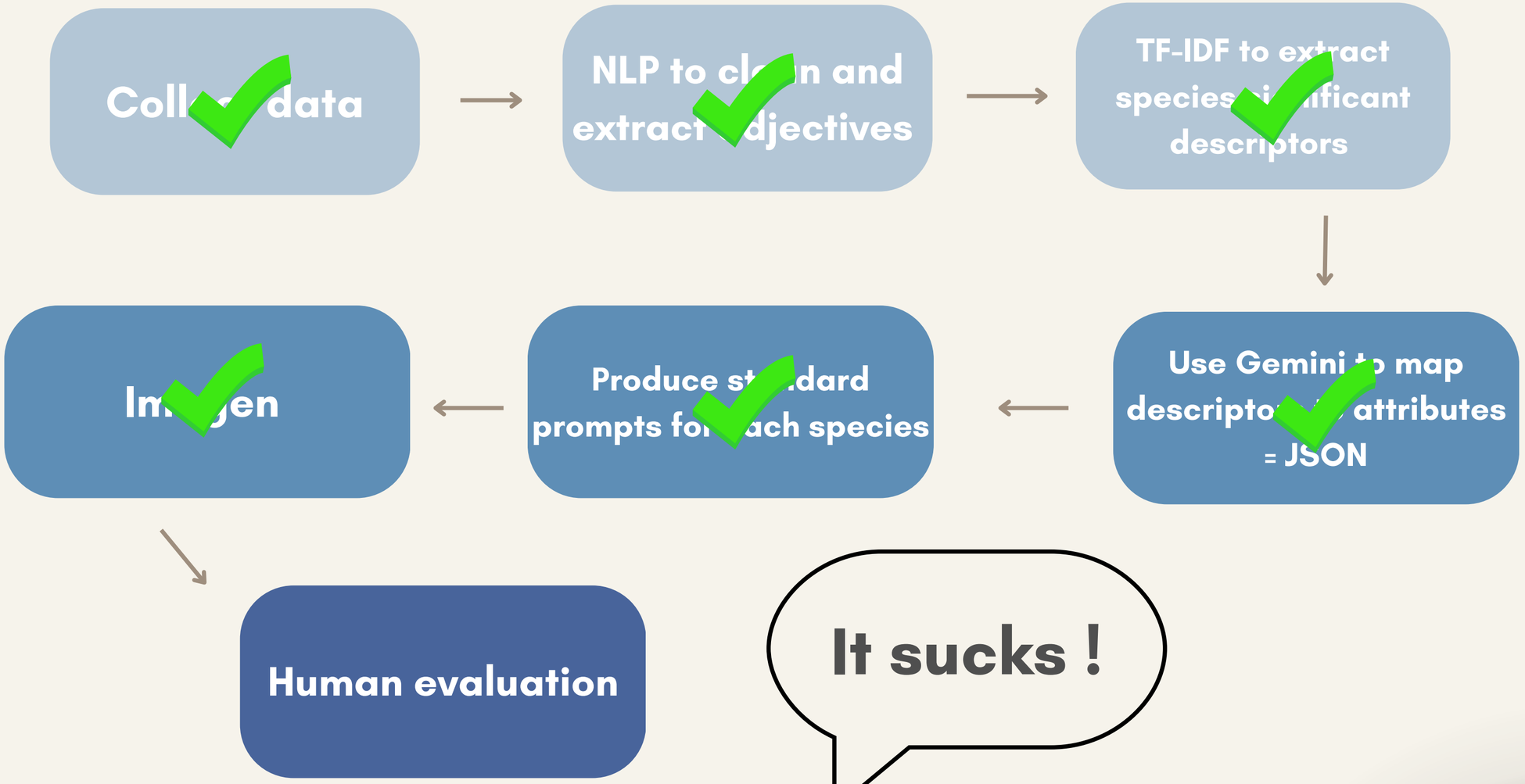
White Sturgeon (Sc)

White Sturgeon (Se)

PROJECT STAGES



PROJECT STAGES



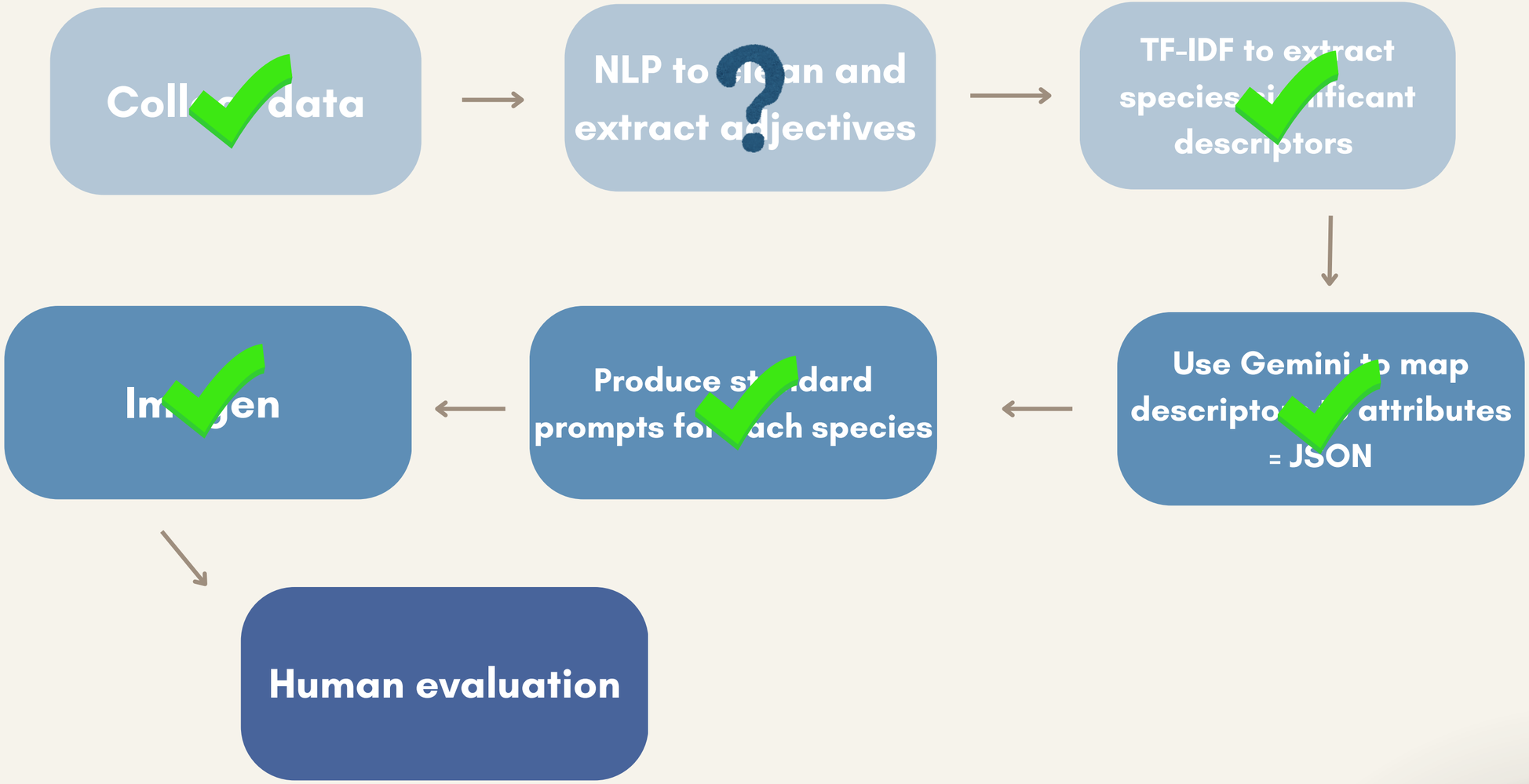
Imagen

White Sturgeon (Sc)

White Sturgeon (Se)

image of a single river fish, colours: white; shape: long; special features: fins; high detail, underwater, soft lighting, naturalistic

PROJECT STAGES

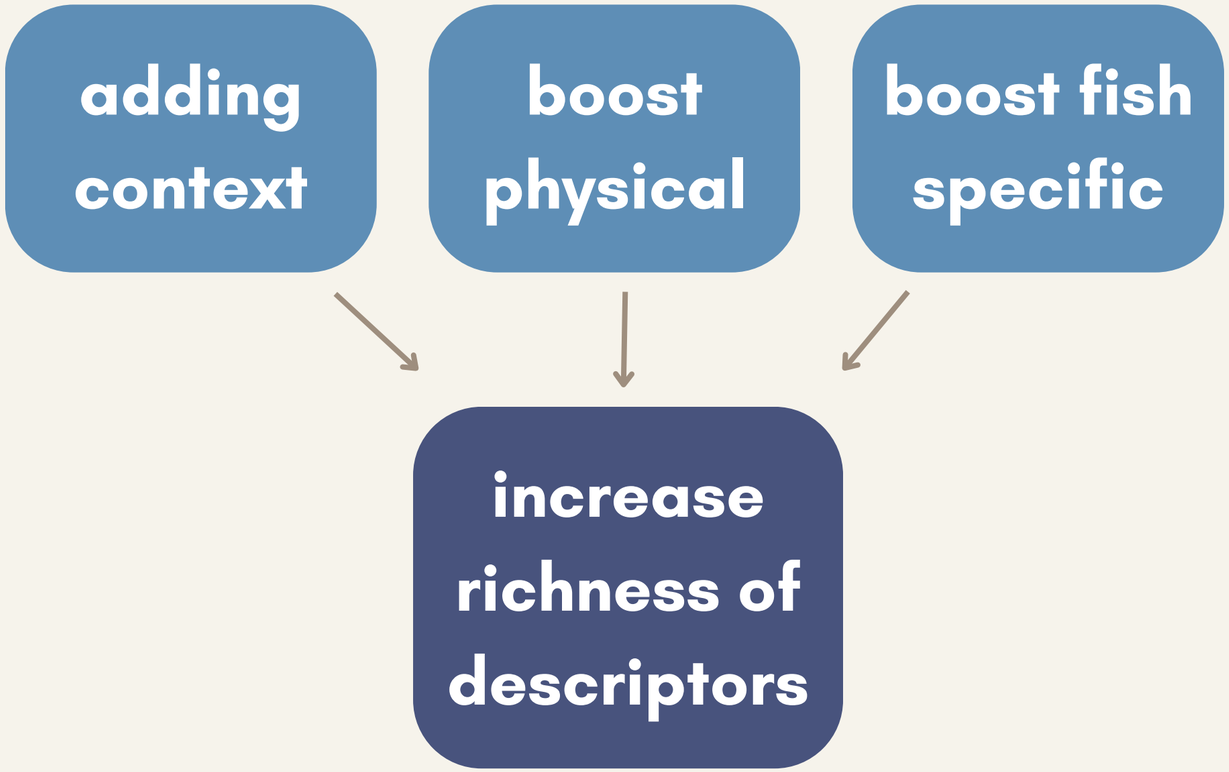


NLP to clean and extract adjectives + TF-IDF

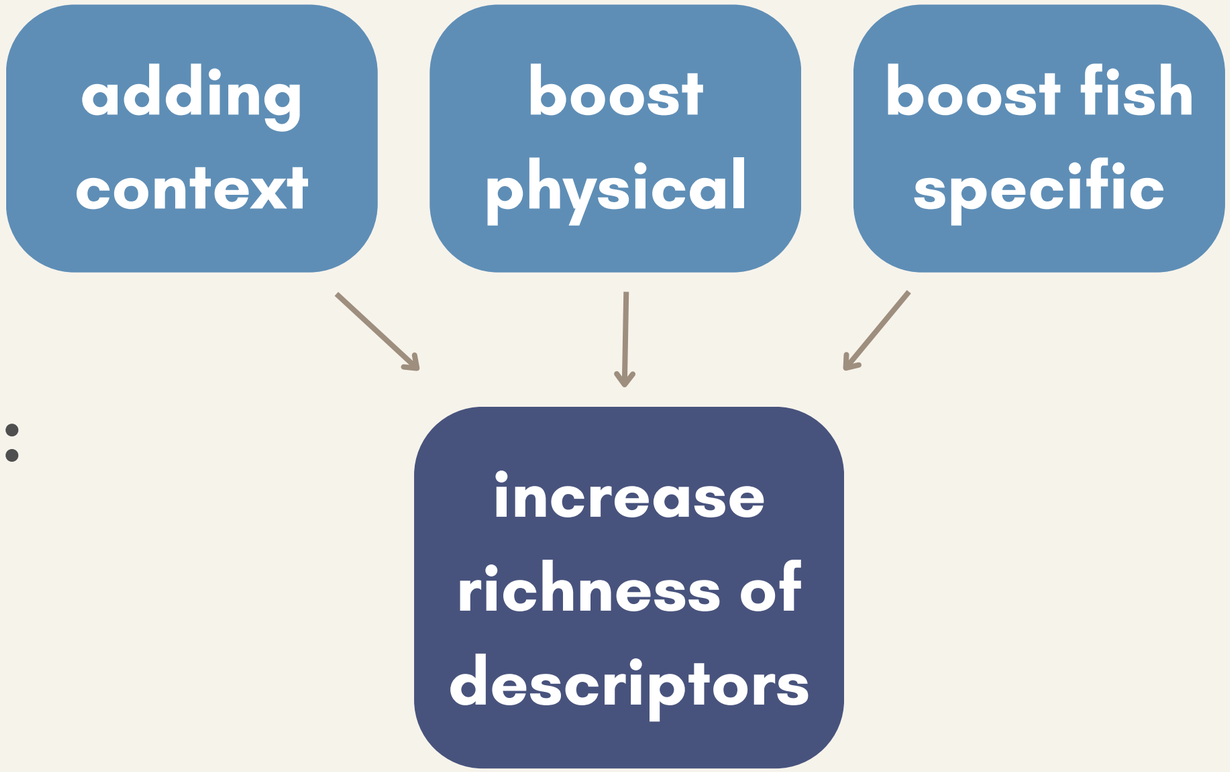
- Adjective + noun pairings, phrases, special nouns
- Boost physical descriptions when scoring TF-IDF to capture distinct features
- Boost domain specific language

NLP to clean and extract adjectives + TF-IDF

- Adjective + noun pairings, phrases, special nouns
- Boost physical descriptions when scoring TF-IDF to capture distinct features
- Boost domain specific language



NLP to clean and extract adjectives + TF-IDF



These changes meant I was able to extract descriptors like:

- ‘wide mouth’
- ‘needle-like teeth’
- ‘elongated body’
- ‘silver scales’

Back on track ..

Use Gemini to map descriptors to attributes = JSON → Produce standard prompts for each species → Imagen

Prompt structure:

f“ {core_instruction} {features} {style}”

Imagen

Feeding NEW prompts to Imagen

Imagen

Feeding NEW prompts to Imagen

White Sturgeon (Se)

White Sturgeon (Sc)

Human evaluation

Goonch catfish

dynamic scene of a strange predatory fish lurking underwater, with a flat, flat bottom form; pointed teeth, eater shark; notable for armor, armor plating; appearing aggressive behavior, aggressive character. Style: photorealistic, high detail on textures and skin, full body

(Se) Accuracy = 3%

Goliath tigerfish

detailed biological illustration of a freshwater predator, with crunching jaws; characterised by poisonous spines, spines; a ferocious predators, fierce predator presence. Style: true to like, anatomical accuracy, dark waters background

(Sc) Accuracy = 63%

Red bellied pacu

illustration of a giant freshwater fish, a body of bright chest, distinct chest hues; rounded profile body; with carnivorous diet; a migratory presence. Style: photorealistic, high detail on textures and skin, full body

(Sc) Accuracy = 40%

Human evaluation

Accuracy Scoring Framework

species	language type	colour	patterns	shape	mouth teeth	special features	Accuracy (%)
example fish1	sensational	2	1	3	5	2	43
example fish1	scientific	4	4	3	4	4	63

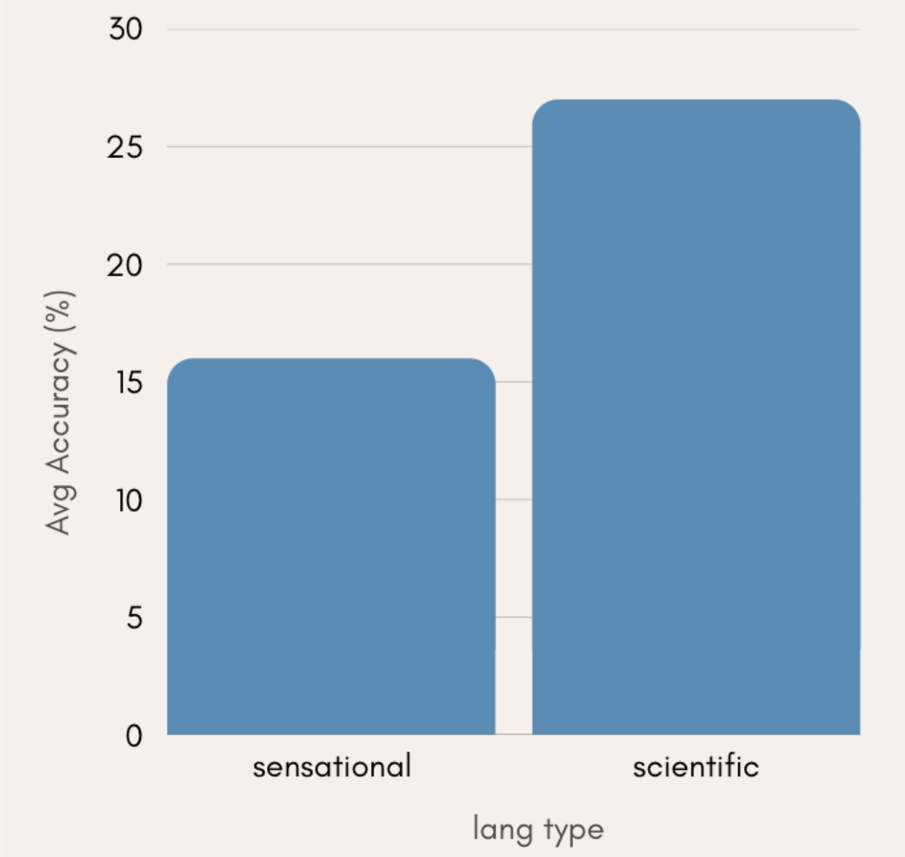
I chose human evaluation as CLIP doesn't really capture species classification tasks well

- **I was the only one to score the accuracies - very subjective**
- **In the future, increase the number of raters just to get more robust accuracy scores**

Human evaluation

Accuracy Scoring Framework

species	language type	colour	patterns	shape	mouth teeth	special features	Accuracy (%)
example fish1	sensational	2	1	3	5	2	43
example fish1	scientific	4	4	3	4	4	63



I chose human evaluation as CLIP doesn't really capture species classification tasks well

- **I was the only one to score the accuracies - very subjective**
- **In the future, increase the number of raters just to get more robust accuracy scores**

Limitations

- The source of the data were sometimes very limited
  - some wiki pages were sparse
  - seek testimonial accounts outside of River Monsters
- Improve NLP and key descriptors extraction
- Imagen can cling on to descriptors like what happened with ‘shark’
- Improve the prompts - maybe add a line to tie it to reality
- Expand and improve the evaluation framework
- Currently a work in progress..

Reflections

- I aimed to use this project an opportunity to learn and challenge myself
- Cleaning is never over
- Importance of asking for help when you need it
- Always good to just see it working first and improve later
- The process can be non linear and that’s okay