

Charting the Conversational Cart: A Voice-Driven Adventure into Creating Virtual Assistants for Grocery Shopping

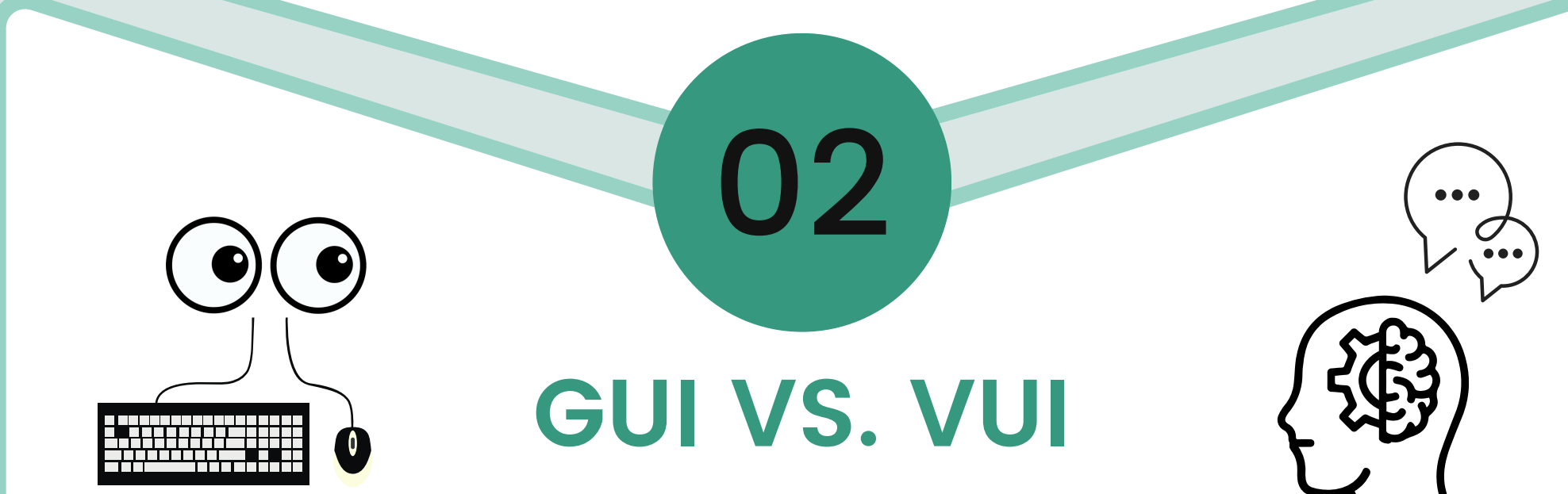
Advised by
 Dr. Sofia Visa
 Department of
 Computer Science



01

RESEARCH INTEREST

- HCI**: This field stresses the importance of human-centered design, the limitation of traditional GUI that drives the development of more intuitive, natural mode of interaction, like touch or voice
- VOICE ASSISTANT**: As a VUI, a voice assistant (VA) is a user-friendly solution to handle automatable tasks, improving daily efficiency
- GROCERY SHOPPING**: The desire for convenience & efficiency when doing grocery shopping aligns well with the capability of a voice assistant



02

GUI VS. VUI

visual-based	voice-based
2-tier interaction	1-tier interaction
hierarchical structure	flow-chart structure
slow input - quick output	quick output - slow output
high cognitive load	low cognitive load
UI design process	voice design process

The optimal user interface is **multi-modal** - leveraging both GUI and VUI

SUMMARY

- Explored the development of human-computer interface, specifically voice user interface (VUI)
- Looked into the core technologies inside a *voice assistant* - a type of VUI, which consists of four parts: (1) *speech recognizer*, (2) *intent recognition engine*, (3) *business logic*, (4) *speech synthesizer*
- Created **Dubee** - an in-app voice assistant for grocery shopping which combines both voice and visual interface to optimize accessibility

#HCI

#VUI

#voice_assistant

#grocery

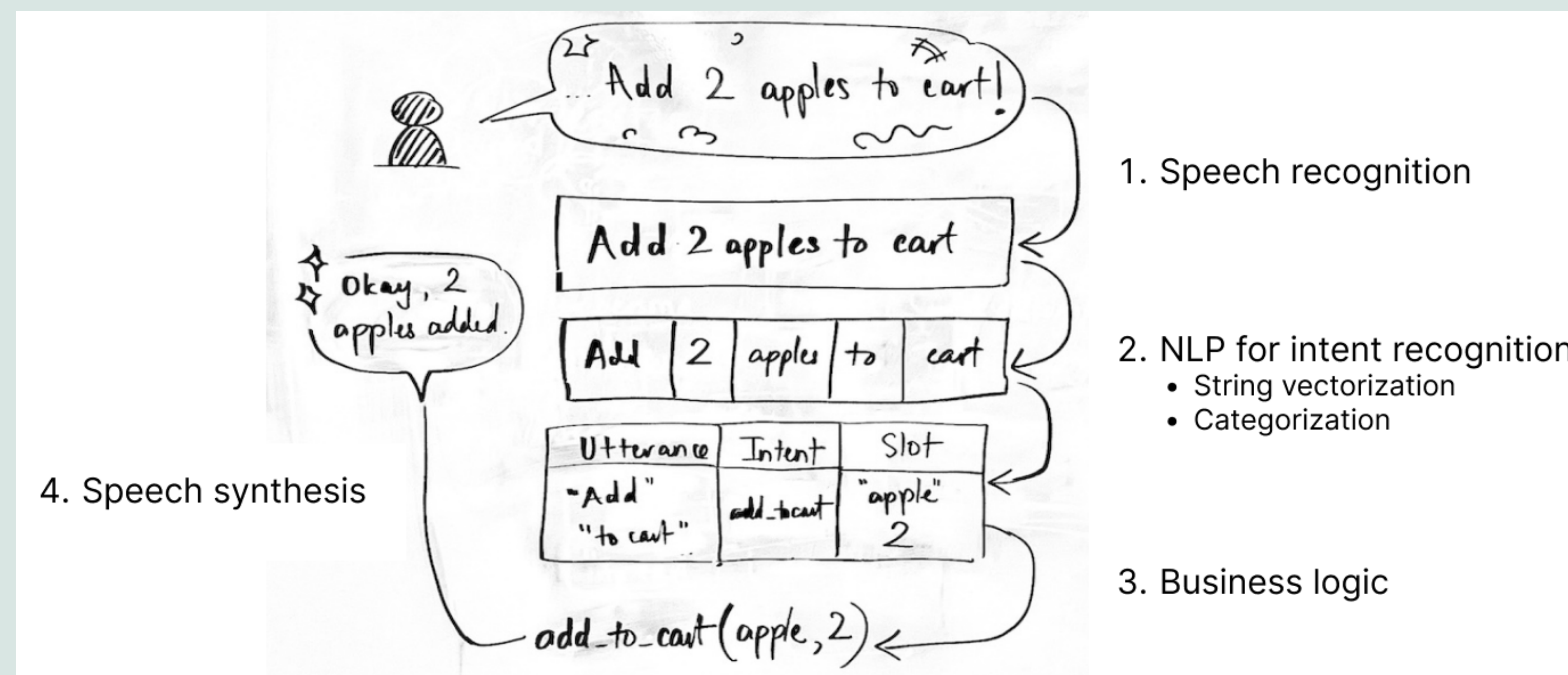


Figure 1: How a VA processes a voice command

```
{
  "intents": [
    {
      "tag": "greeting",
      "patterns": ["Hey", "Hello", "Hi", "Hi there"],
      "responses": ["Hello there, I'm Dubee"]
    },
    {
      "tag": "add_to_cart",
      "patterns": ["Add to cart", "Purchase"],
      "responses": ["Okay, item added"]
    }
  ]
}
```

Figure 2: Intent recognition JSON file

03 INSIDE A VOICE ASSISTANT

- Speech Recognition** (speech-to-text, STT) uses machine learning models to identify and transcribe different accents, dialects, and speech patterns, from there convert spoken language to texts
- Natural Language Processing (NLP)** enables machines to understand, interpret, and respond to natural language input from users. In voice assistants, NLP algorithms are used to create an intent recognition engine.
- Business Logic via Hooks** refers to the UX engineering of the assistant based on the unique business scope of the product
- Speech Synthesis** (text-to-speech, TTS) generates speech from texts. It has evolved over the year in terms of sophistication and voice varieties and currently holds an important role in digital accessibility

INTENT RECOGNITION [FIGURE 2]

- Identify what the user intends to achieve through a query or message.
- Common methods are based using dictionaries, rule-based; statistical methods; and deep-learning.
- Rule-based model:** used in early chatbots and virtual assistants; uses conditionals to identify keywords/phrases; these entities trigger predefined responses. This is applied in our software by using a JSON file [Figure 2]

Example from Figure 1:
 1. User requests "Add two apples to cart"
 2. Model is trained to recognize "add", "two", "apples", and "cart" as entities, evaluating the intent as "add_to_cart"
 3. Model is programmed to match this intent to a function inside the logic system
 4. Provides a predefined response after the function is carried out, "Okay, item added"

04 CREATING DUBEE \doo-bee\

Design Process: Understand - Explore - Materialize

We created a user persona, a user journey map, and some sample dialogues to aid with the human aspect of our software. This is also the standard procedure in designing voice-based interactions.

System Design

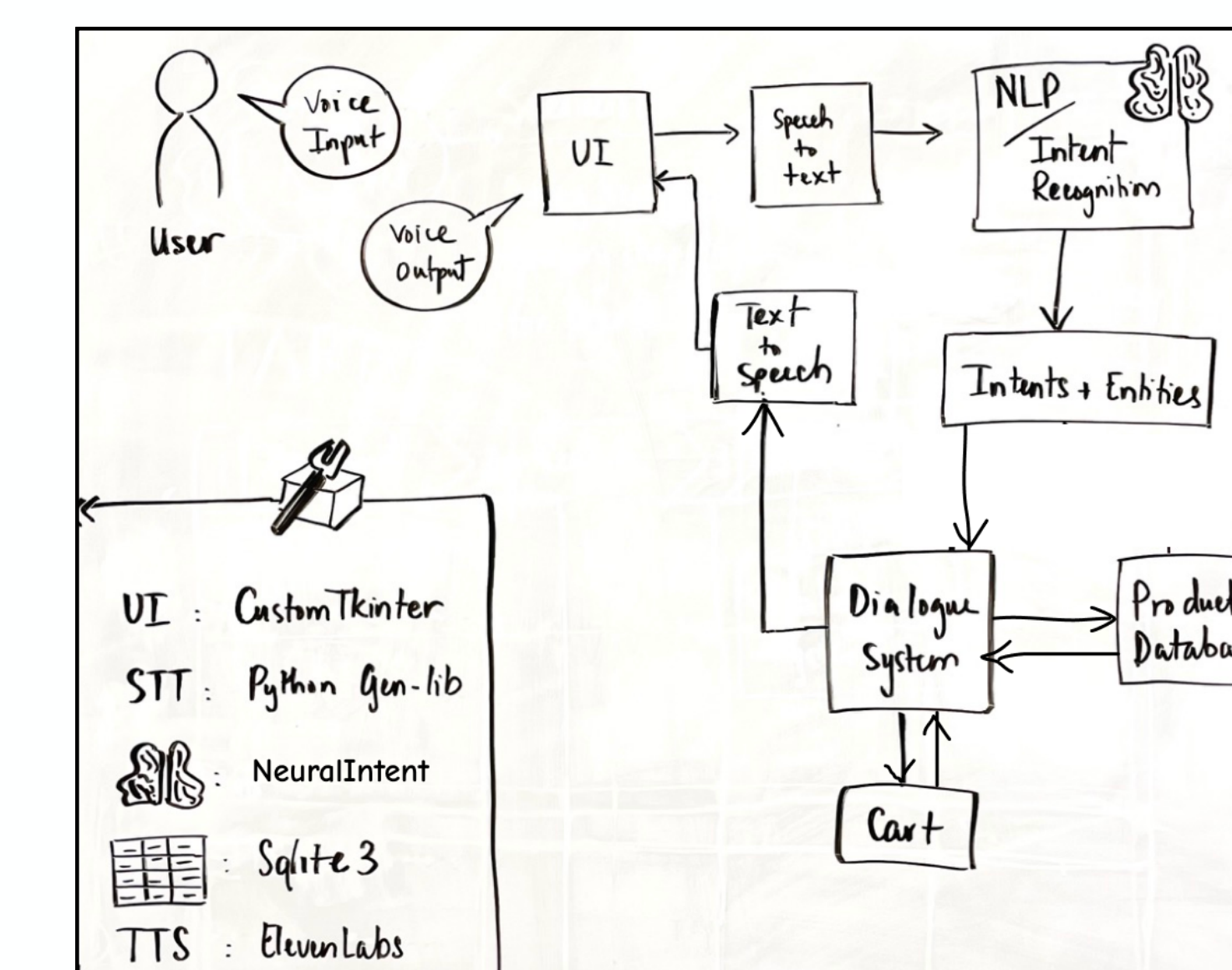
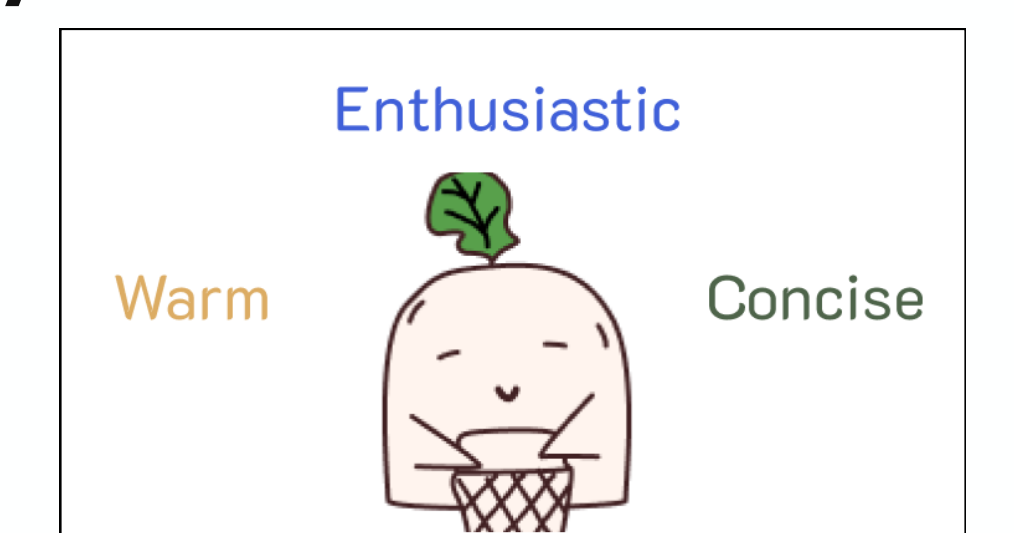


Figure 3: Dubee's system design

Coding tools

- Python language
- CustomTkinter UI library
- Speech Recognition library
- ElevenLabs Text Synthesis API
- SQLite3 for database

System Persona



05

CONCLUSION & OUTLOOK

- Dubee is a modest, from-the-ground-up prototype for the future of designing accessible grocery shopping experiences
- A successful voice assistant fosters trust between human-computer through its unique persona, the efficiency and visibility of system's state, and natural conversation understanding
- VUI holds the key to the future of interaction: safer navigation aid inside vehicles, vision-impairment-friendly museum guide, quick & responsive assistant inside hospitals, etc.

Works Cited

- Section 1's HCI logo: bit.ly/hci_logo
- Section 2: bit.ly/voice_ui
- Section 3's intent recognition: bit.ly/rec_intent
- Section 4's voice design process: bit.ly/voicedesignprocess