



JARVIS

Content



1

INTRODUCTION

2

How Jarvis
work

3

What's
Different

4

Future
Improvement

1

INTRODUCTION



"We have developed an advanced AI assistant that integrates NLP, generative text and image capabilities, and speech recognition to provide intuitive, on-demand support."

Content



1

INTRODUCTION

2

How Jarvis
work

3

What's
Different

4

Future
Improvement

2

How Jarvis work



Combines between

NLP

GEN
Text
&
Code

GEN
Image

Speech
Rec.

NLP

Intent Prediction



Datasets

Snips built in intents

Amazon Science massive

Custom Data

Preprocessing

Text Correction (Spelling & Grammar)

Lower & Contractions

Tokenization

Lemmatization

Model

Data Processing Layer

Label encoding with scikit-learn's LabelEncoder

Text tokenization using BERT tokenizer

Data sampling mechanism for training efficiency

Model Architecture

Pre-trained BERT base (uncased) backbone

Sequence classification head for intent prediction

Training System

AdamW optimizer with fine-tuning learning rate (3e-5)

Batch processing with attention masking

Epoch-based training loop with loss tracking

Data Flow

1.Text → Tokenization → BERT Encoding → Classification Head → Intent Prediction

2.Training data → Sampling → Batching → Model → Loss Calculation → Backpropagation

NLP

Conversation

uses the Gemini AI API to enable interactive conversations.



key components:

Core Components

API Integration Layer

Interfaces with Google's Gemini AI API
Handles authentication and request formatting
Manages API responses and error handling

Session Management System

Maintains multiple independent chat sessions
Provides session creation, switching, and deletion
Implements persistent storage through file saving/loading



User Interface Layer

Text-based menu system
Command processing and input validation
Response formatting and display

Data Persistence Layer

JSON-based file storage
Session import/export functionality
Timestamp and metadata tracking

Key Functions

`call_gemini()`: Handles API communication
`show_history()`: Displays conversation logs
`save_sessions_to_file()`: Persists data
`load_sessions_from_file()`: Retrieves saved sessions
`main()`: Orchestrates the application flow

Data Flow

User input → Input processing → API request → Response handling → Display
Session data → JSON serialization → File storage → File loading → Session restoration

NLP

Samples



Intent Prediction

Text:OPEN CAMERA -----> Predicted intent: open application
Text:SEND EMAIL TO BOSS -----> Predicted intent: email send email
Text:PLAY NEXT SONG -----> Predicted intent: next track
Text:ما هو الوقت الآن -----> Predicted intent: calendar set
Text:WHAT'S THE WEATHER IN CAIRO -----> Predicted intent: weather query
Text:SET A TIMER FOR 10 MINUTES -----> Predicted intent: alarm set
Text:SEND LOCATION TO MY FRIEND -----> Predicted intent: share current location

Conversation

You

how can i sleep like a normal human being

02:14 AM

Jarvis

Okay, so you're having trouble sleeping. That's really common! Let's try to get you sleeping more like a "normal human being" (though remember, everyone's sleep needs are a little different!). Here's a comprehensive approach, breaking it down into manageable steps:

2

How Jarvis work



Combines between

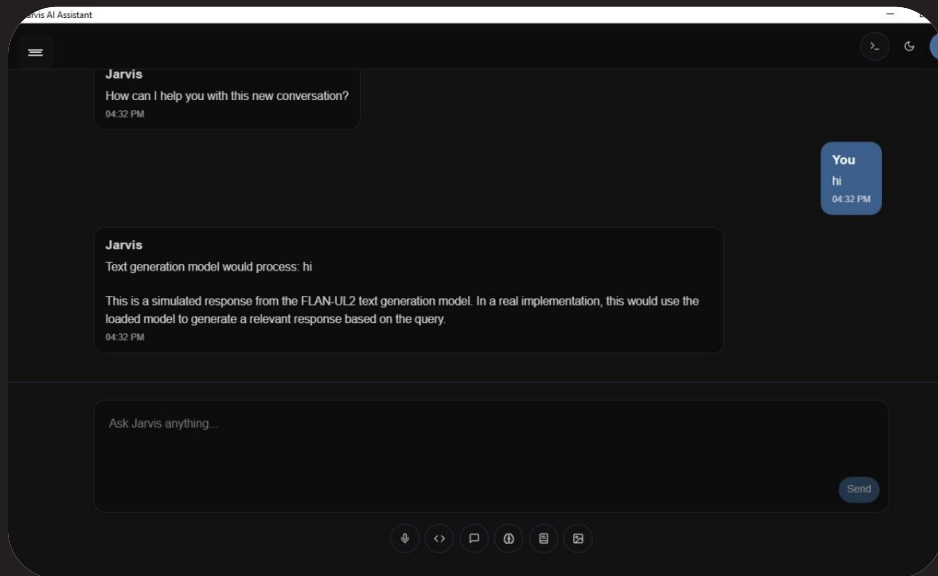
NLP

GEN
Text
&
Code

GEN
Image

Speech
Rec.

Generative Text & Code



A comprehensive AI assistant framework featuring multiple model architectures, training pipelines, and dataset processing capabilities. This project includes code generation, text generation, and CNN-enhanced text generation models.



Gemini-2.0-Flash for quick responses

Generative Text & Code

End-to-end fine-tuning pipeline for large language models



****FLAN-UL2 (20B) – Text**

Generation**

- * Encoder-decoder architecture ideal for complex generation tasks
- * Instruction-tuned with diverse tasks for better generalization
- * Built on UL2: supports multiple pretraining objectives
- * Strong in dialogue and persona consistency tasks

Two primary tasks:

- 1 Code Generation with DeepSeek-Coder (6.7B)
- 2 Text Generation with FLAN-UL2 (20B)

Stages:

Data Processing → Feature Extraction →
PEFT Training → Evaluation → Visualization

****Model Selection****

****DeepSeek-Coder (6.7B) – Code Generation****

- * Pretrained specifically on code (multi-language: Python, JS, Java, etc.)
- * Strong instruction-following baseline for code tasks
- * Good performance-to-size trade-off (trainable on consumer GPUs)
- * Open weights enable unrestricted fine-tuning

Generative Text & Code



****Dataset Overview – Code Generation****

****CodeSearchNet: 2M examples****

→ Paired code and natural language from 6 languages (Python, Java, JS, etc.)

****Code Alpaca: 20,022 examples****

→ Instruction-response format tailored for coding

****Magocoder-OSS-Instruct: 75,197 examples****

→ High-quality filtered instruction dataset for practical development tasks

Generative Text & Code

****Dataset Overview – Code Generation****



****MBPP: 1,401 examples****

→ **Basic programming problems with structured inputs and test cases (used for training & eval)**

****HumanEval: 164 examples****

→ **Evaluation benchmark with complex programming challenges (also included in fine-tuning)**

****CodeParrot: 5.17M examples (~50GB)****

→ **Cleaned and deduplicated real-world Python files from GitHub**

→ **Used to teach natural coding styles and patterns**

Generative Text & Code



****Dataset Overview – Text Generation****

****OpenAssistant: 88,838 examples****

→ High-quality human dialogues, multi-turn, instruction-following focus

****GPTeacher: 82,260 examples****

→ Educational instructions with deep explanatory responses

****The Pile: 825 GiB of text****

→ Broad domain coverage: literature, coding, math, philosophy, etc.

→ Boosts general knowledge and generation diversity

****PersonaChat: 10,906 examples****

→ Character-based conversations with consistent personas

→ Useful for role-playing and personality alignment

****WritingPrompts: 303,358 examples****

→ Creative writing stories paired with prompts

→ Enhances narrative coherence and storytelling capability

Generative Text & Code



****Dataset Processing & Training Pipeline****

****Processing Highlights:****

- * Language-specific handling (esp. for CodeSearchNet)**
- * Deduplication and quality filtering to reduce noise**
- * Caching and batched tokenization for memory efficiency**

****Training Techniques:****

- * LoRa + 4-bit quantization (PEFT)**
- * Gradient accumulation, mixed precision (FP16)**
- * Deep Speed Zero + gradient checkpointing for memory savings**
- * Dataset weighting to balance training across tasks**

Generative Text & Code



****Evaluation Strategy****

****Functional Correctness****

- * Computes pass@k for HumanEval and MBPP
- * Executes generated code in sandboxed environments
- * Validates outputs against official test cases

****Code Quality****

- * Evaluates cyclomatic complexity
- * Measures code readability and maintainability
- * Style compliance via linting tools (e.g., flake8, black)

****Runtime Performance****

- * Measures execution time
- * Tracks peak memory usage
- * Compares efficiency relative to reference implementations

****Evaluation Methods****

- * ``ModelEvaluator.evaluate_humaneval()``: Runs HumanEval benchmark
- * ``ModelEvaluator.evaluate_mbpp()``: Runs MBPP benchmark
- * ``ModelEvaluator.evaluate_code_generation()``: Evaluates code generation on custom prompts
- * ``ModelEvaluator.evaluate_code_quality()``: Assesses quality metrics (lint, complexity)
- * ``ModelEvaluator.evaluate_runtime()``: Measures runtime and memory usage
- **Semantic Evaluation****
 - * Compares semantic similarity to ground truth
 - * Analyzes logical soundness of generated code
 - * Evaluates problem-solving strategy beyond surface-level match

Generative Text & Code

Samples



Jarvis

Okay, so you're having trouble sleeping. That's really common! Let's try to get you sleeping more like a "normal human being" (though remember, everyone's sleep needs are a little different!). Here's a comprehensive approach, breaking it down into manageable steps:

****1. Establish a Consistent Sleep Schedule:****

- * ****Go to bed and wake up at the same time every day,**** even on weekends. This helps regulate your body's natural sleep-wake cycle (circadian rhythm). Even if you had a terrible night, try to stick to your wake-up time. This is SUPER important.
- * ****Expose yourself to bright light in the morning.**** This helps solidify your circadian rhythm. Open your curtains, go outside, or use a light therapy lamp.

Jarvis

Okay, so you're having trouble sleeping. That's really common! Let's try to get you sleeping more like a "normal human being" (though remember, everyone's sleep needs are a little different!). Here's a comprehensive approach, breaking it down into manageable steps:

****1. Establish a Consistent Sleep Schedule:****

- * ****Go to bed and wake up at the same time every day,**** even on weekends. This helps regulate your body's natural sleep-wake cycle (circadian rhythm). Even if you had a terrible night, try to stick to your wake-up time. This is SUPER important.
- * ****Expose yourself to bright light in the morning.**** This helps solidify your circadian rhythm. Open your curtains, go outside, or use a light therapy lamp.

2

How Jarvis work



Combines between

NLP

GEN
Text
&
Code

GEN
Image

Speech
Rec.

Generative Image



The Image Generation module of Jarvis AI Assistant implements cutting-edge diffusion models and GANs to create high-quality images from text descriptions. The system features advanced fine-tuning capabilities, allowing for personalization and customization of the generated images.

Generative Image



Stable Diffusion

Turning Words Into Worlds

**What if you could sketch your thoughts with
only your words**

Generative Image



What Is Stable Diffusion?

Definition

“Stable Diffusion is a deep learning model that generates high-quality images from text prompts using diffusion techniques.”

Text Prompt → Encoder → Noise →
Denoising Process → Final Image

Generative Image

The Problem It Solves



**Humans imagine faster than they can draw
or render.**

Challenges:

Designers waste hours on mockups

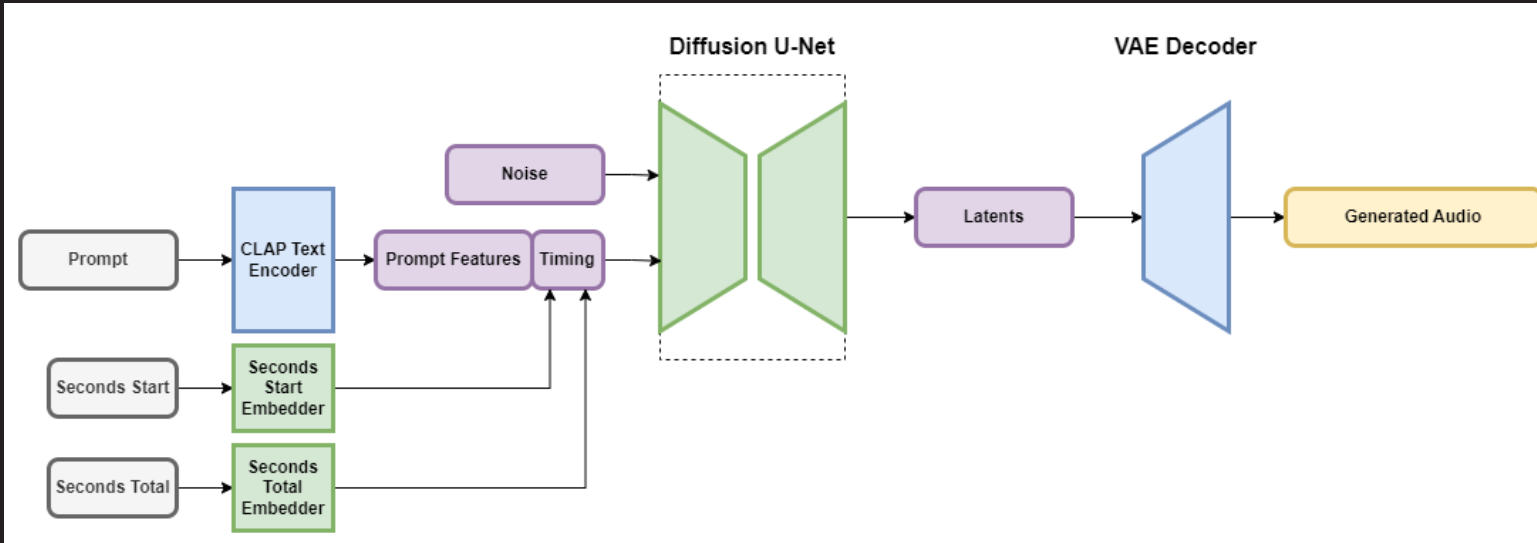
Writers lack visuals for their stories

Marketers need fast, scalable content

Developers struggle with placeholder visuals

Problem Line: “Imagination is limitless. Time isn’t.”

Generative Image



The Magic Behind the Model
Technical Breakdown
(simplified):

Latent diffusion process

CLIP text encoder understands prompts

U-Net denoises step-by-step until an image is born

Powered by billions of image-text pairs



Generative Image



Why It Stands Out

Feature	Stable Diffusion	DALL·E	Midjourney
-----	-----	-----	-----
Open-source	Yes	No	No
Customizable	Easily fine-tuned	Limited	No
Offline Capable	Yes	No	No
Speed	Fast	Slower	Fast
Community	Huge open community	Research-focused	Artists only

Samples

Unleashing Visual Power
"One prompt. Infinite imagination."



"blond woman high quality"

Showcase a gallery:

"A golden dog"



"A high quality cat"

"A black and white photo of a horse in space"



Samples



**Our Custom-Built Stable Diffusion
DreamBooth Fine-Tuning**

**I trained the model on my own images using
DreamBooth — now it can generate realistic, stylized
images of me in any scene or context**



Samples

LoRA Fine-Tuning
fine-tuning LoRA adapters to specialize the model in
specific art styles like vintage posters and pixel art



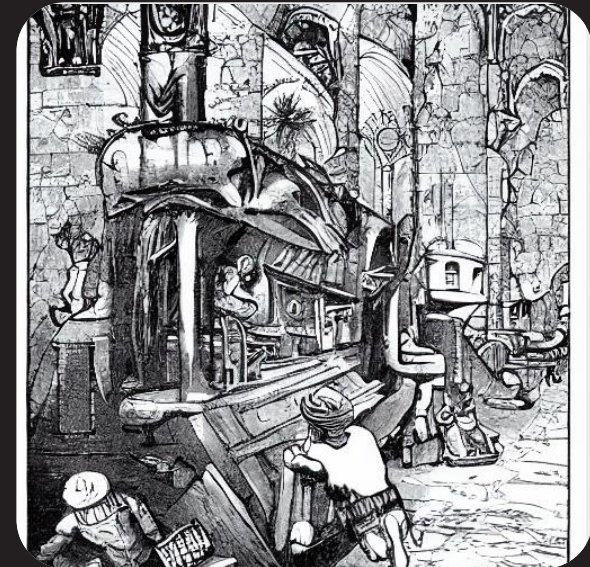
“an old book illustration of a
mountain”

“A man in a street attacked
by a snake”



“An old man is sitting on
the blooming branch of a
cherry tree and scatters
ashes from a basket”

“A very old market”



Samples

LoRA Fine-Tuning
fine-tuning LoRA adapters to specialize the model in
specific art styles like vintage posters and pixel art



“armored mech archer”



“Crimson mech knight,
glowing eyes, ruined city”



“Neon samurai”

“Alien bounty hunter”



2

How Jarvis work



Combines between

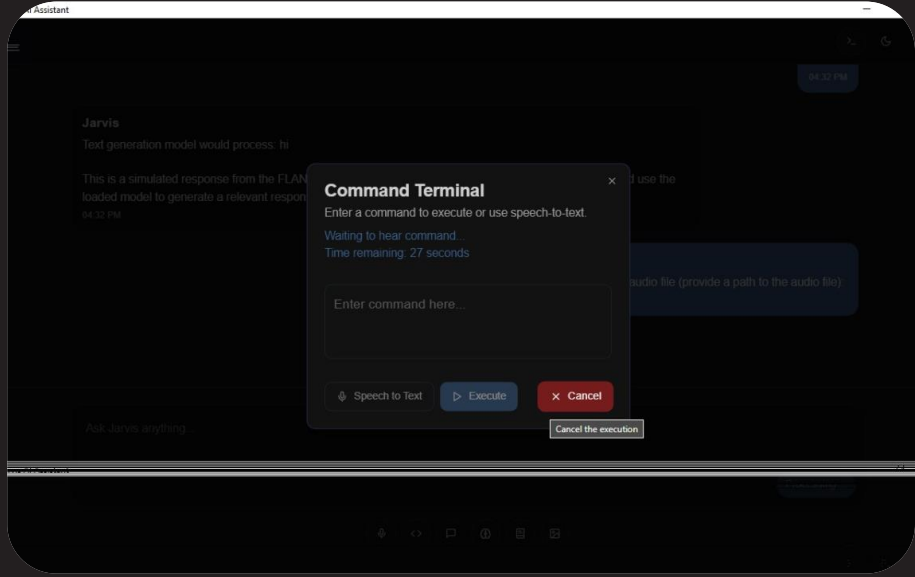
NLP

GEN
Text
&
Code

GEN
Image

Speech
Rec.

Speech Recognition



This project implements a real-time speech command recognition system using TensorFlow and a pre-trained deep learning model. The system can identify 30 common spoken commands from a user's microphone input, including numbers, directions, and basic words.

Support for 30 different speech commands

Speech Recognition



Fine tuning Whisper-small on Egyptian Arabic Dialect dataset

Why whisper-small ? Small resources with very good word error

Dataset Summary for Presentation

Egyptian Arabic Dialect Automatic Speech Recognition (ASR)

This dataset is optimized for training/fine-tuning ASR models (e.g., Whisper) and is hosted on Hugging Face Hub.

Key Features:

Content: 16 hours of Egyptian Arabic speech from diverse YouTube broadcasts (e.g., talk shows, news).

Transcription: Manually annotated by 4 transcribers per clip to address dialectal variations (no formal orthographic rules).

Structure:

Audio: 16kHz samples, segmented into 30-second clips (ideal for memory-efficient training).

Text: Egyptian dialect transcriptions paired with each audio segment.

Speech Recognition

Fine tuning Whisper-small on Egyptian Arabic Dialect dataset



Dataset Creation:

Sourced from MGB-3 Challenge: 80+ YouTube programs, each with the first 12 minutes transcribed.

Processed into 30-second segments after filtering unavailable/deleted videos

Use Cases:

Train/fine-tune ASR models for Egyptian Arabic, a widely spoken dialect with limited standardized resources.

Benchmark model performance on dialect-rich, real-world speech.

Accessibility:

Publicly available on Hugging Face Hub, pre-cleaned and ready for use.



Why It Matters:

Addresses the challenge of dialectal diversity in Arabic ASR, enabling robust voice technology for Egyptian speakers.

Source: MGB-3 Challenge, YouTube broadcasts.

2

How Jarvis work



Combines between

NLP

GEN
Text
&
Code

GEN
Image

Speech
Rec.

Content



1

INTRODUCTION

2

How Jarvis
work

3

What's
Different

4

Future
Improvement

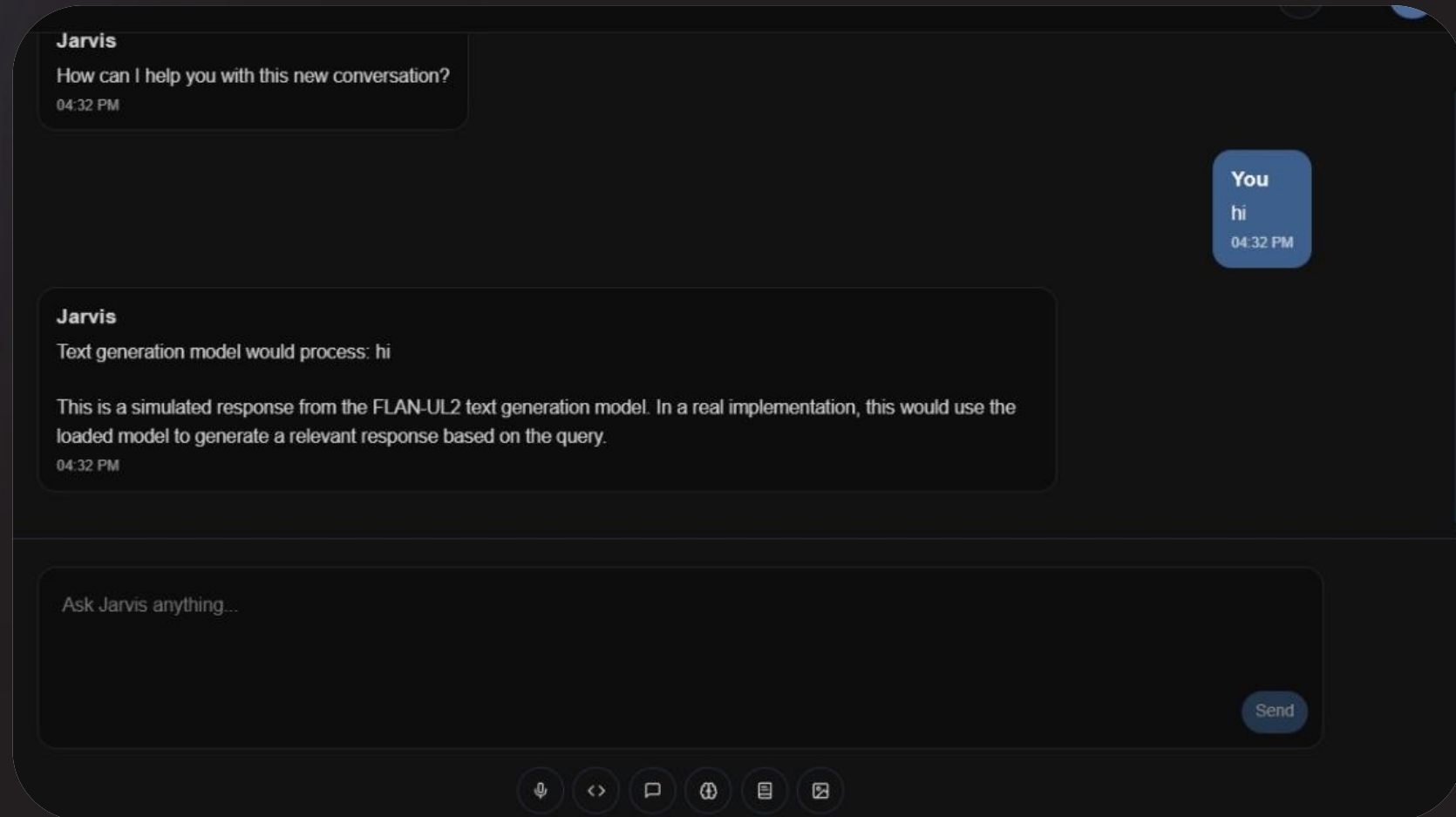
3



“

What's Different Between
Jarvis & Siri & GPT

3 The UI Preview



3

What's Different Between Jarvis & Siri & GPT



Predict
Intent

Access to
Your APPS

Makes
Perfect
Images

Code
Generation

Good
Conversation

Voice
Assistant

Simple
UI

Content



1

INTRODUCTION

2

How Jarvis
work

3

What's
Different

4

Future
Improvement

4

Future Improvement



Health
Care

More
Options

Makes
APP &
Website

For all fields
of work

Generative
Video

All Languages
Voice Assist

More
Friendly

Content



1

INTRODUCTION

2

How Jarvis
work

3

What's
Different

4

Future
Improvement

“

“Innovation distinguishes
between a leader and a
follower.”

— Steve Jobs

Ahmed Gamal



Amr Khaled

Nada Ayman

Hamza Osama