

Aurora: Assistente de Voz Inteligente para Automação Local e Aumento de Produtividade

Descrição Geral da Ideia

O projeto consiste em desenvolver um assistente de voz inteligente e local, baseado em interfaces não convencionais (InCs), que utiliza real-time speech-to-text com wakeword para interagir com o usuário. O sistema reconhece comandos de voz, processa as intenções do usuário por meio de um Large Language Model (LLM) e executa ações específicas, como buscar informações em registros de atividades diárias (via OpenRecall) ou controlar um navegador (via browser-use). A resposta é entregue em áudio, criando um diálogo fluido e natural.

O diferencial do Aurora está em sua operação totalmente local e anônima, garantindo privacidade, e na integração de ferramentas de código aberto para expandir funcionalidades sem substituir interfaces tradicionais. Por exemplo, o usuário poderá perguntar: “O que eu pesquisei sobre interfaces não convencionais às 14h?”, e o assistente usará o OpenRecall (que captura screenshots periódicos e indexa semanticamente) para recuperar a informação. Simultaneamente, comandos como “Abra meu e-mail e pesquise a última mensagem do professor” serão executados via automação no navegador.

Materiais e Métodos

Wakeword Detection:

Biblioteca: Porcupine (leve e offline) para detecção da palavra de ativação (ex: “Aurora”).

Justificativa: Baixa latência e compatibilidade com processamento local.

Speech-to-Text (STT) em Tempo Real:

Modelo: Whisper da OpenAI (versão tiny ou base para operação local via Transformers).

Pipeline: Captura de áudio via microfone e conversão contínua para texto.

Large Language Model (LLM) e Gerenciamento de Ferramentas:

Modelo: Llama 3 (7B ou 13B, quantizado para eficiência) ou Mistral 7B.

Framework: LangChain ou LlamaIndex para orquestrar chamadas de ferramentas externas:

OpenRecall: Integração via API local para busca semântica (RAG) em registros de atividades (ex: screenshots indexados).

Browser Automation: Uso do browser-use para controle de navegador (ex: abrir abas, preencher formulários com o uso de simples prompts).

Lógica: O LLM decide se responde diretamente ou aciona uma ferramenta com base no contexto.

Text-to-Speech (TTS):

Ferramenta: Piper (síntese de voz offline, com vozes customizáveis).

Infraestrutura:

Hardware: Computador com GPU dedicada (para LLM) ou CPU moderna (via quantização).

Sistema Operacional: Linux (Ubuntu) para compatibilidade com bibliotecas open-source.

Fluxo de Funcionalidade

O usuário ativa o assistente com a wakeword ("Aurora").

O áudio subsequente é convertido para texto via Whisper.

O texto é processado pelo LLM, que decide se:

Responde diretamente (ex: perguntas gerais).

Chama o OpenRecall para buscar informações em registros históricos para compor uma resposta mais completa.

Executa ações no navegador via browser-use (ex: abrir sites, clicar elementos).

A resposta final é sintetizada em áudio via Piper e reproduzida para o usuário.

Segurança e Privacidade

Todos os dados (áudio, texto, screenshots) são processados localmente.

Justificativa para Interfaces Não Convencionais

O Aurora utiliza a voz como interface primária, rompendo com a dependência de telas e teclados, e introduzindo interações mais intuitivas e acessíveis. A combinação de wakeword, processamento de linguagem natural e automação contextual cria uma experiência que amplia a produtividade sem substituir interfaces existentes — por exemplo, um usuário pode ditar um comando enquanto continua digitando. Além disso, o uso de ferramentas como o OpenRecall (que transforma screenshots em banco de dados semântico) redefine como informações visuais são recuperadas, integrando visão computacional e linguagem natural de forma não convencional. Por fim, a operação local e anônima desafia o paradigma atual de assistentes baseados em nuvem, alinhando-se a demandas éticas crescentes. Este projeto demonstra como InCs podem ser amplificadoras (não disruptivas) do fluxo de trabalho, oferecendo novas camadas de interação que coexistem com métodos tradicionais.