University of Nottingham
UK | CHINA | MALAYSIA

Department of Electrical and Electronic Engineering

Bangbang 2.0:

Multimodal Perception and AI Assistance for Inclusive Telepresence



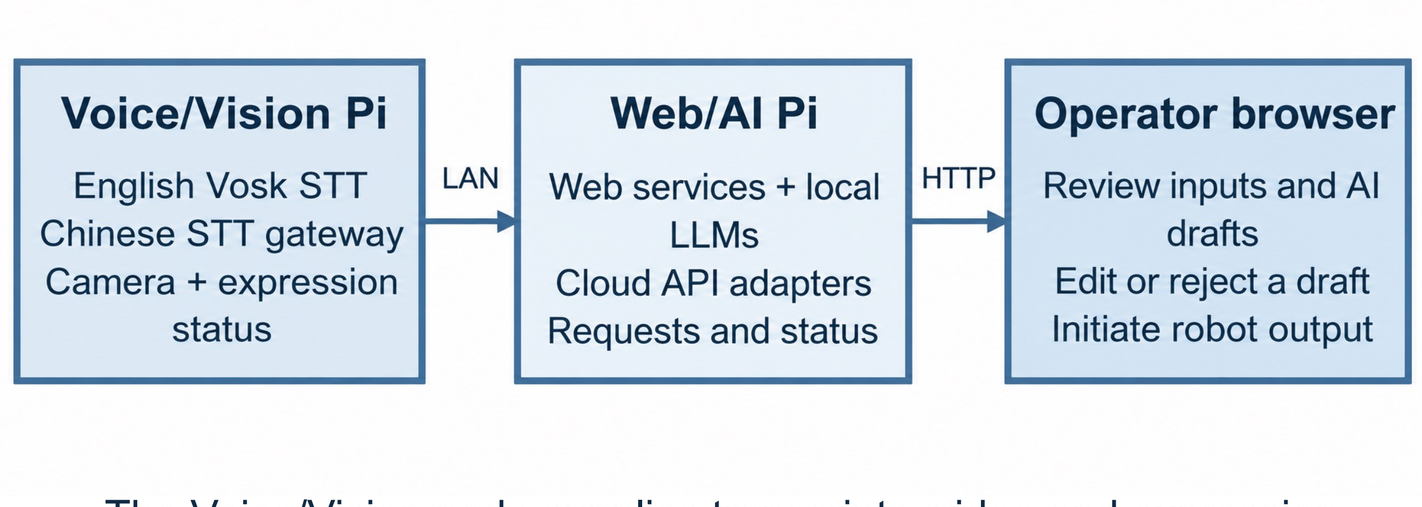
1 Motivation and problem

Bangbang is a desktop avatar for inclusive remote service work. We combine bilingual speech and facial-expression cues with optional LLM drafts, so an operator can interpret visitors and decide what the robot communicates. The two studies cover perception and operator assistance.

2 Hardware and shared platform

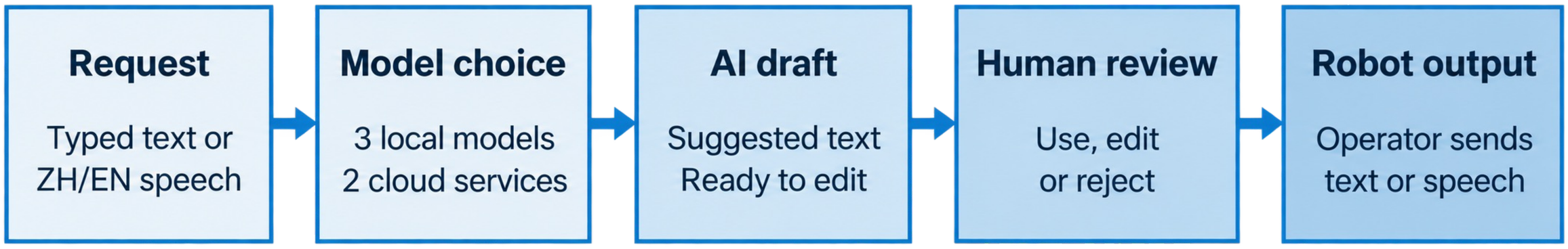


Figure 1. Shared robot and teleoperation setup



The Voice/Vision node supplies transcripts, video and expression status. The Web/AI node serves the operator interface and routes selected model requests.

3 Operator workflow



Perception context

Live video and facial-expression labels support the operator. The person may answer directly or select another endpoint.

Local and cloud assistance

Local: Qwen2.5-0.5B, DeepSeek-R1-Distill-Qwen-1.5B and Qwen3-4B.
Cloud: DeepSeek chat and Gemini 2.5 Flash.

Operator authority

Choose an endpoint, check the draft and decide what reaches the visitor. Visible errors allow manual retry or a different path.

4 Perception results

CNN and facial action-unit (AU) classification

Measure	CNN	AU rules
Expression classes	7	6
Pi 5 CPU usage*	Approx. 86%	Approx. 47%
Neutral calibration	None	30 frames
Interpretation	Learned features	Named action units

Perception prototype: approx. 30 FPS.
*CPU observations use the paper's four-core basis.
AU = facial action unit.

Chinese and English speech

English: offline Vosk. Chinese: Baidu Cloud STT.
Speech produces an editable transcript.

Evidence limits

Functional checks establish integration. User benefits remain untested.
Robot speech is team-confirmed without retained logs.

5 Integrated interaction

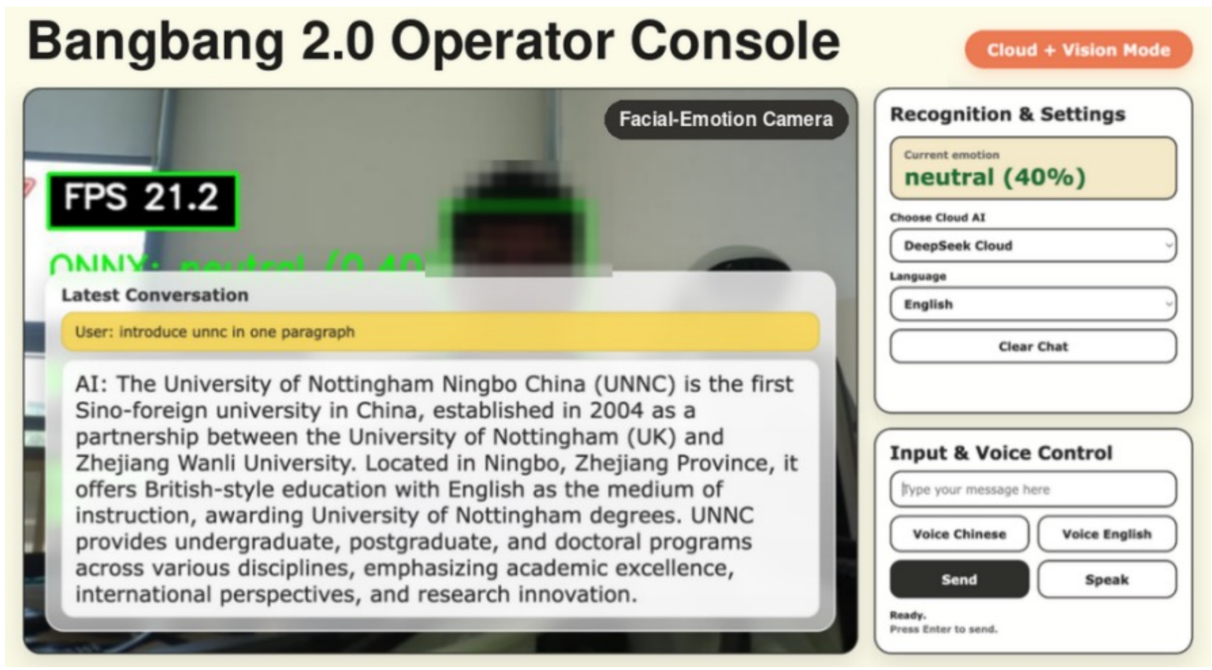


Figure 2. The console combines video, expression status and voice controls.

6 Future work

Test guidance accuracy and workload with intended users. Compare direct answers, fixed AI and manual model choice under shared bilingual tasks.

Changshuai Liu, Yanjie Zhou, Siheng Ye and Yuechen Cao
Supervisor: David Chieng Co-supervisor: Younghoon Chang