Shibo Liu

Undergraduate Student · Reliable On-device AI Systems & Local Agents

spencerliu@stu.ncst.edu.cn
nihildigit.dev
github.com/NihilDigit
ORCID 0009-0004-9426-3713

Summary

Undergraduate student working on reliable on-device AI across mobile perception, NPU inference, and local agents. Focus on latency-critical, privacy-preserving, and accessibility-oriented deployments.

Selected Work

ANVIL - Accelerator-Native Video Interpolation ***2026***

*Sole-author manuscript under review at **IEEE TCSVT** **arXiv:2603.26835***

- Built a 30-to-60 fps video interpolation system under mobile NPU, INT8, and end-to-end playback latency constraints.
- Reused H.264 decoder motion vectors as a deployment-friendly motion prior, replacing learned optical flow and GridSample-heavy flow refinement with prealignment plus a convolutional residual network; additive skips cut latency 17-26% in A/B tests.
- Forked mpv-android into a real playback harness and ran the full CPU/GPU/NPU pipeline on Xiph 1080p sequences: 12.8 ms 1080p INT8 NPU inference, 28.4 ms median end-to-end latency over 54,623 consecutive frame pairs, 94.9% within the 33.3 ms budget.

TongXing - On-device Assistive Agent ***2025-2026***

4C2026 Provincial Second Prize; national round in progress

- Built the on-device safety loop for blind and low-vision mobility: detection → depth → segmentation → tracking → distance/TTC → warning priority, all on a single Android phone.
- Deployed INT8 vision models on Hexagon NPU and a Q4-quantized Qwen3-1.7B on Adreno GPU via llama.cpp/OpenCL, with offline speech/control on CPU. Safety perception averaged 35.16 ms over 1,636 walking frames, under the 66.7 ms budget for a 15 Hz warning loop.
- Closed the digital-accessibility gap in touch-first apps by routing navigation, ride-hailing, transit, weather, and vision tasks through a voice loop backed by MCP tools and Android accessibility actions, while keeping the safety path local and isolated from assistant tasks.

RAFNet - Dense Classroom Behavior Recognition ***2025-2026***

*Under review at **The Visual Computer***

- Third author; designed evaluation and ablations for gated fusion over **GroupRec (ICCV 2023)** relation context and ConvNeXt appearance features.
- Reached 63.08 ± 0.40 Macro F1 on the self-built NCST Classroom dataset (+2.79 over ConvNeXt-only) and 96.76 cross-dataset Macro F1 on SCB3-U.

Open Source

- **Animeko** / **pikpak-kotlin**: merged PikPak cloud offline-download into open-ani/animeko, an 18k-star app, with streaming, BitTorrent fallback, credential handling, and a Maven-published KMP SDK.
- **waybar-ai-usage**: maintains an AUR-packaged Waybar quota widget for coding agents; 40+ stars and 8 accepted external contributions.

Education

B.S. ***2023-2027***
Intelligence Science and Technology

North China University of Science and Technology, Tangshan, China

- CET-4: 548.

Awards

- **4C2026**: Provincial Second Prize; national round in progress.
- **4C2025**: 3rd Prize, Hebei provincial round.
- **APMCM 2024**: Second Prize.

Skills

Efficient edge ML: mobile NPU/GPU heterogeneous pipelines, INT8 quantization, latency-budgeted inference, media/playback integration, profiling and benchmarking.

Local agent systems: on-device LLM deployment, MCP tool-use workflows, Android accessibility automation, offline-first design.

Human-centered deployment: interaction design, accessibility-oriented UX, task-oriented interfaces, user-facing reliability.

Links

- **TongXing Showcase**
- **ANVIL Showcase**
- **ANVIL arXiv**
- ANVIL: **Code** / **Player**
- **RAFNet OSS Repo**