

CONDENSED & SHIPPABLE BUILD PLAN

Goal: A jaw-dropping, magical voice agent in ≤ 30 days

Total tasks: 8 (instead of 27)

Everything else is deferred to v2.

```
text           X ⌂ Copy

voice-agent/
└── frontend/
    ├── index.html
    └── static/
        ├── style.css
        ├── app.js          → becomes tiny coordinator
        ├── audio.js
        ├── avatar.js       ← NEW
        ├── theme.js         ← NEW
        └── settings.js     ← NEW
    └── manifest.json    ← NEW (PWA)

└── server/
    ├── main.py
    ├── config.py        ← NEW (simple)
    └── session.py       ← NEW (tiny in-memory)

├── sw.js            ← NEW (PWA service worker)
├── config.yaml
├── requirements.txt
└── launch.sh
```

The 8 Tasks That Actually Matter (do them in this exact order)

Task 1 – Refactor & Persist Settings (1–2 days)

Create three tiny modules from existing app.js:

- settings.js → localStorage + defaults (theme, volume, voiceSpeed, reducedMotion)
- theme.js → applyTheme(), cycleTheme()
- avatar.js → setState(state), setExpression(expr)

Update index.html → <script type="module" src="static/app.js"></script>

Result: Settings survive refresh, code is sane.

Task 2 – Make the Avatar Feel Alive (2–3 days)

Pure CSS + tiny JS. All of these at once:

Idle:

- Subtle breathing (scale 0.98 → 1.0 over 4s)
- Random blink every 3–7s
- Very subtle floating (translateY ±2px)
- Occasional tiny head tilt (±4deg)

Listening state:

- Eyes 5% wider, no blinking, slight forward lean

Thinking state:

- Eyes up-right, slower breathing

Speaking state:

- Mouth open/close synced to audio volume (use Web Audio analyser, 4 lines of code)

Expressions (overlay classes):

- .happy, .curious, .thinking, .calm

Respect @media (prefers-reduced-motion)

This single task is responsible for 70% of the “holy shit” reaction.

Task 3 – Instant & Reliable Barge-in (<200 ms) (1–2 days)

Backend (main.py):

- During TTS playback → temporarily lower VAD threshold to 0.35
- On any speech detection → immediately send { "type": "interrupt" }

Frontend (audio.js):

- On "interrupt" → `AudioContext.close()`, clear queue, stop everything instantly
- Avatar instantly switches to listening state

Test with real interruptions — must feel like Siri/Alexa.

Task 4 – Simple Config + In-Memory Session (1 day)

Create server/config.py (pydantic + config.yaml override)

Create server/session.py — just a dict[websocket] → list of last 10 messages

Update LLM prompt to include:

text

Copy

Previous messages (most recent first):

User: ...

Assistant: ...

...

No disk, no timeouts, no complexity yet.

Task 5 – Beautiful Slide-Out Settings Panel (2–3 days)

Right-side glass panel (mobile = full-screen):

- Theme swatches (larger grid)
- Volume slider (speaker)
- Voice speed slider 0.75–1.5× (sent to backend on change)
- Reduced motion toggle
- Clear conversation button
- Close with Esc or swipe left

All changes instant, saved via settings.js

Task 6 – Mobile-First Responsive + Touch Magic (2 days)

- Flex/grid layout → works perfectly down to 320px
- Orb becomes 72px on mobile
- Tap orb = start/stop recording (no need to hold)
- Long-press orb = classic push-to-talk
- Swipe right edge → open settings
- Add viewport meta, safe-area insets, no double-tap zoom

Test on your actual phone daily.

Task 7 – PWA + 4 Keyboard Shortcuts (1 day)

Create manifest.json + 192px/512px icons (just use the avatar face)

Create minimal sw.js (cache HTML/CSS/JS, runtime caching for everything else)

Add these four shortcuts only:

- Space → push-to-talk
- Esc → stop recording / stop speaking
- Ctrl + / → toggle settings
- ? → show tiny shortcut toast

Done. Now installable on phone & desktop.

Task 8 – Final Polish & Ship (1–2 days)

- Add typing indicator (three bouncing dots) when streaming
- Auto-scroll with “jump to bottom” floating button
- Empty state: “Say hello” with avatar wave animation
- Favicon + apple-touch-icon
- Update README with 15-second GIF of it working

New MISTAKES.md (only 7 lines)

Markdown

X ⌂ Copy

1. Don't add tests before people love it
2. Don't build complex session manager before week 4
3. Don't animate more than breathing + blink + expressions
4. Don't make barge-in slower than 200 ms
5. Don't ship without testing on a real phone
6. Don't add features – remove them
7. Don't forget: magic = avatar + instant barge-in

Success = You have this in ≤30 days:

- Installable on home screen
- Gorgeous breathing avatar that reacts perfectly
- Never talks over you
- Works flawlessly on phone and desktop
- Friends say “this feels better than ChatGPT voice”

Everything else (tests, advanced config, virtual scroll, 17 shortcuts, offline page, etc.) is v2.

Want me to write the exact 8 prompt-ready tasks (with full file paths and code snippets) so you can just paste them one by one?

Say “GO” and I’ll give you the copy-paste version that gets you shipped in 30 days.

↳ Expand Task 2 avatar details

↳ Similar open-source voice agents

↳ Add code snippets to tasks