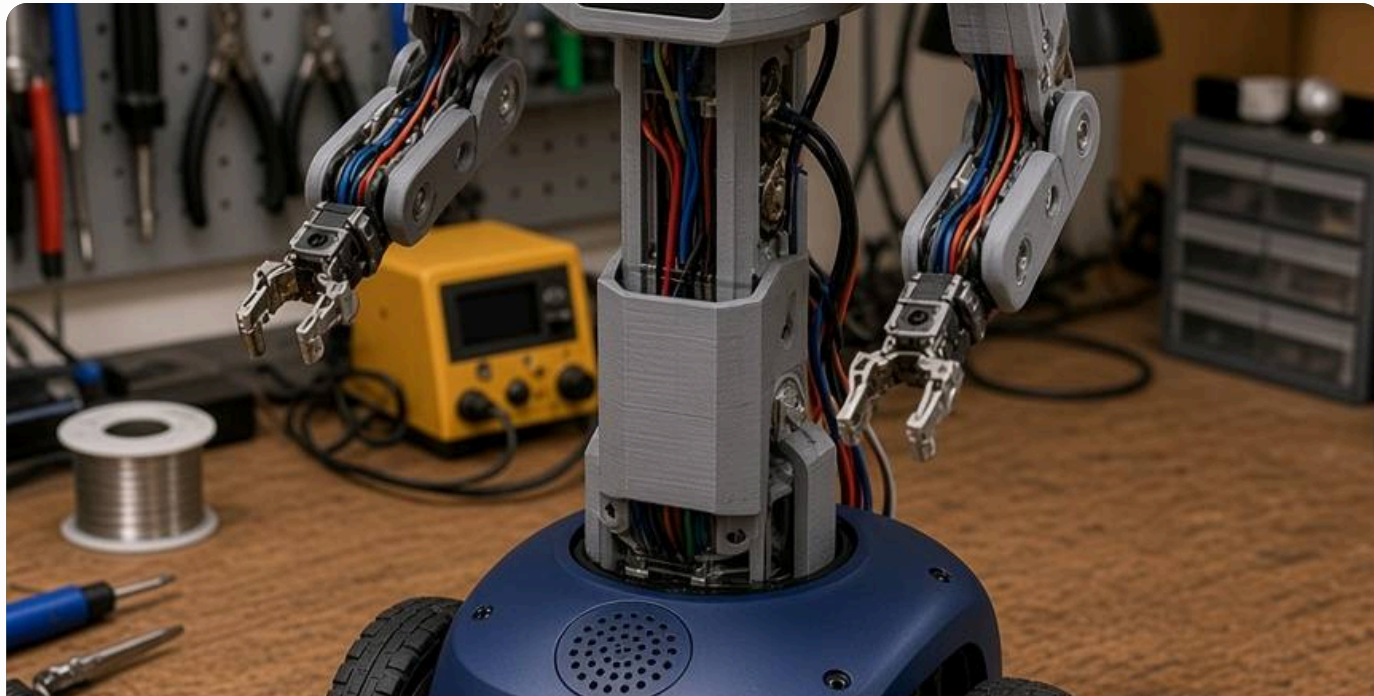


Grok Bot: Your AI, With Hands

A build-it-yourself robot with Grok's voice, vision and reasoning. Minimal about \$615, nicer about \$1,265.



COPPER GATOR ROBOTICS

Grok Bot: Your AI, With Hands

Grok's voice and vision in a robot you build yourself. It rolls, looks around, talks back, and picks things up.

Concept render (AI-generated, not a photo): the full Grok Bot, with a blue wheeled base, a gray torso, two arms with grippers, and a camera head, on a workbench

Why you'll love it

It talks and sees

Ask it what it sees and it answers out loud. Grok handles the voice, camera and reasoning.

It moves and grabs

Two-wheel drive, shoulders, a pan/tilt head, and soft printed gripper fingers.

Never blacks out

A spine-mounted battery plus a backup pack keep its brain running through battery swaps.

Built by you

A Pi 5 brain, an ESP32 for motors, and parts you print and solder yourself. You can upgrade any of it.

Pick your build



Concept render (AI-generated): left, an open-frame minimal rover with a Raspberry Pi, a camera, and ultrasonic sensors; right, an enclosed blue base with a lidar puck on top. Bases only, no torso or arms shown.

MINIMAL

~\$615

Wheels, camera, voice, spine battery, backup pack. The fastest path to a talking robot.

NICER

~\$1,265

Lidar, encoder motors, a beamforming mic, LiFePO4 power and a face display.

STRUCTURAL PACK OPTION

~\$1,460

The Nicer build with a load-bearing 4S LiFePO4 spine pack for long runtimes.

Staged Build: pay as you go

Buy it in phases, cheapest first. Every stage works on its own before you spend another dollar, and Stage 1 runs on the Raspberry Pi 4B and 10-inch touchscreen you already own. Prices are estimates as of 23 Oct 2026 (Pi 5 2GB verified at \$77.50 list).

STAGE 1

Talking head on your desk

\$70 Running total \$70

Unlocks: Grok voice conversation plus camera vision, with the touchscreen as its face and control panel.

- Pi Camera Module 3, \$25
- USB mic, \$15
- MAX98357A I2S amp + 3W speaker, \$10
- 64 GB microSD (dedicated robot card), \$10
- Pi 4 USB-C 5V 3A supply if needed, \$8
- Printed desk stand, about \$2 of PETG

STAGE 2

Wheels: it comes to you

\$271 Running total \$341

Unlocks: Drives room to room on voice command, avoids obstacles, reports what it sees.

- Chassis, 2x JGB37-520 encoder motors, wheels, \$56
- TB6612FNG driver + ESP32-S3 real-time board, \$21
- 3x HC-SR04P + MPU-6050 IMU, \$13
- 3S 5000 mAh LiPo + BMS + thermal break, \$51
- Fuse, switch, XT60, 5V 5A buck, \$30

STAGE 3

Torso, head and shoulders

\$124 Running total \$465

Unlocks: A 45-55 cm body with pan/tilt head that looks at you, and shoulders that gesture.

- 2020 extrusion torso + printed shell, \$40
- Pan MG996R on lazy-susan + MG90S tilt + slip ring, \$29
- 2x MG996R shoulders + PCA9685, \$24
- 6V 10A servo UBEC, \$15
- Spine battery cradle + hardware, \$16

Reuses: Your Pi 4B and 10-inch touchscreen (saves about \$68 Pi + \$60-90 screen)

- 6S-capable AC/DC balance charger, \$55 (shared with every drone)
- Wiring + PETG, \$45

Reuses: Stage 1 head; the Pi 4 rides on the base

Reuses: Stage 2 battery, ESP32, Pi 4

STAGE 4

Pi 5 brain + keep-alive power

\$161 Running total \$626

Unlocks: Hot-swap batteries with no reboot, faster vision. Frees your Pi 4 to become the full-time drone ground station.

- Raspberry Pi 5 2GB, \$77.50 list
- Active Cooler, \$5
- Geekworm X1200 UPS + 2x Molicel P28A, \$60
- XT90-S anti-spark pair + VBAT sense, \$8
- 64 GB microSD, \$10

Reuses: Everything above; the Pi 4 moves to drone duty

STAGE 5 (OPTIONAL)

Nicer upgrades, one at a time

\$640 Running total \$1,266

Unlocks: Lidar mapping, far-field mic, stronger motors, LiFePO4 power, outdoor-ready shell.

- reSpeaker XVF3800 mic array, +\$46
- LD19 lidar + 3x VL53L1X ToF, +\$117
- Pololu encoder motors, wheels, Cytron driver, +\$135
- Bioenno LiFePO4 pack + INA226, +\$77
- BNO085 IMU, Feetech bus servos, +\$95
- ASA shell, gaskets, wiring, E-stop, fan, +\$170

ADD-ON: CELLULAR

+\$90

SIM7600G-H dongle. Buy only once the robot leaves Wi-Fi range.

ADD-ON: STRUCTURAL PACK

+\$195

4S A123 pouch spine after Stage 5. Long runtimes, skip on a tight budget.

Stage	Cost	Running total	Unlocks
1 Talking head	\$70	\$70	Voice + vision on your Pi 4 and touchscreen
2 Wheels	\$271	\$341	Drives to you, avoids obstacles
3 Torso + head	\$124	\$465	Pan/tilt head, shoulders
4 Pi 5 + keep-alive	\$161	\$626	No-reboot swaps; Pi 4 freed for drones
5 Nicer (optional)	\$640	\$1,266	Lidar, beamforming mic, LiFePO4

Shared, not double-counted: the Pi 4B and touchscreen (robot until Stage 4, then the drone ground station; until then swap microSD cards to switch roles), and the 6S balance charger bought in Stage 2 also charges the GPS drone and Scout packs.

What you can build with it

Home helper

Fetches small items, patrols rooms, and tells you what it saw.

Patio sentry

Rolls the porch and describes anything unusual.

Learning lab

Hands-on robotics with a real AI brain: code, wiring, 3D printing.

Print the parts: pick your printer



Image: Bambu Lab. Bambu Lab X2D 3D printer, official product photo

RECOMMENDED: BAMBU X2D

\$899

Two nozzles on one head and a heated chamber. Prints PETG brackets and TPU gripper fingers in one job, and nylon too.



Image: Bambu Lab. Bambu Lab P1S 3D printer, official photo; shown without the AMS unit that comes with the Combo

THE CHEVY 350 OPTION: P1S COMBO

~\$549

A proven single-nozzle workhorse with the AMS. Prints PETG just as well; print TPU fingers as separate jobs.

Filament at a glance

PETG

Brackets, motor mounts
Stiff and heat-tolerant: the default material.

PLA

Cheap, easy prints
Good for prototypes, but it softens in a hot car or sun.

TPU

Gripper fingers, dampers
Rubbery and grippy, and it absorbs shock.

Nylon

High-strength parts
Very tough. Print it in the X2D's heated chamber.

Talk to it

Say its name

A wake word runs right on the robot, so nothing is sent anywhere until you call it.

Real conversation

Grok's realtime voice answers in about a second, and you can interrupt it mid-sentence.

Works offline, slowly

If the internet drops, a small local brain still hears and answers, just more slowly and more simply.



Concept render (AI-generated): head and chest close-up. The round 4-mic array sits flat on top of the head, away from the motors; the speaker grille is on the chest panel below, so the mics are not right next to the speaker.

Voice adds about \$10 to the minimal build and about \$15 to the nicer one (the mic and speaker were already in both parts lists; this covers the better mic and the mounting). Cloud conversation costs about \$0.08 a minute.

Stay connected: optional cellular

Texts you when it matters

A tip-over, a safety stop, a low battery or getting stuck: it sends Max a text, even with no Wi-Fi.

Keeps talking off Wi-Fi

If Wi-Fi drops, Grok's voice can keep going over LTE, and you can call the robot to talk to it remotely.

Strictly optional

About +\$90 on the minimal build or +\$120 on the nicer one, plus a plan from about \$2-15 a month. Not included in the prices above.



Concept render (AI-generated): an LTE modem HAT on the Pi with a nano SIM inserted, coax pigtails to two flat LTE antennas and a square GPS patch, and a clear printed window panel behind. Generic board, not an exact product.



Image: Waveshare. SIM7600X 4G HAT for Raspberry Pi (SIM7600G-H variant recommended for the nicer build), with MAIN, AUX and GNSS antenna connectors, USB and audio jack; antennas not shown

See what it sees: video over cellular (optional)

Snapshots and clips

When something happens, it sends a photo or a short clip that Max can open on his phone.

Live view on request

Tap to watch live, with about half a second of delay on a good signal. It only streams when asked.

No new hardware

Uses the same LTE modem and camera. The software is free; the main cost is a bigger data plan.



Concept render (AI-generated): a phone showing the robot's live camera view over LTE. Illustrative only, not a screenshot of a real app.

Hot-swap without a reboot

Brain stays awake

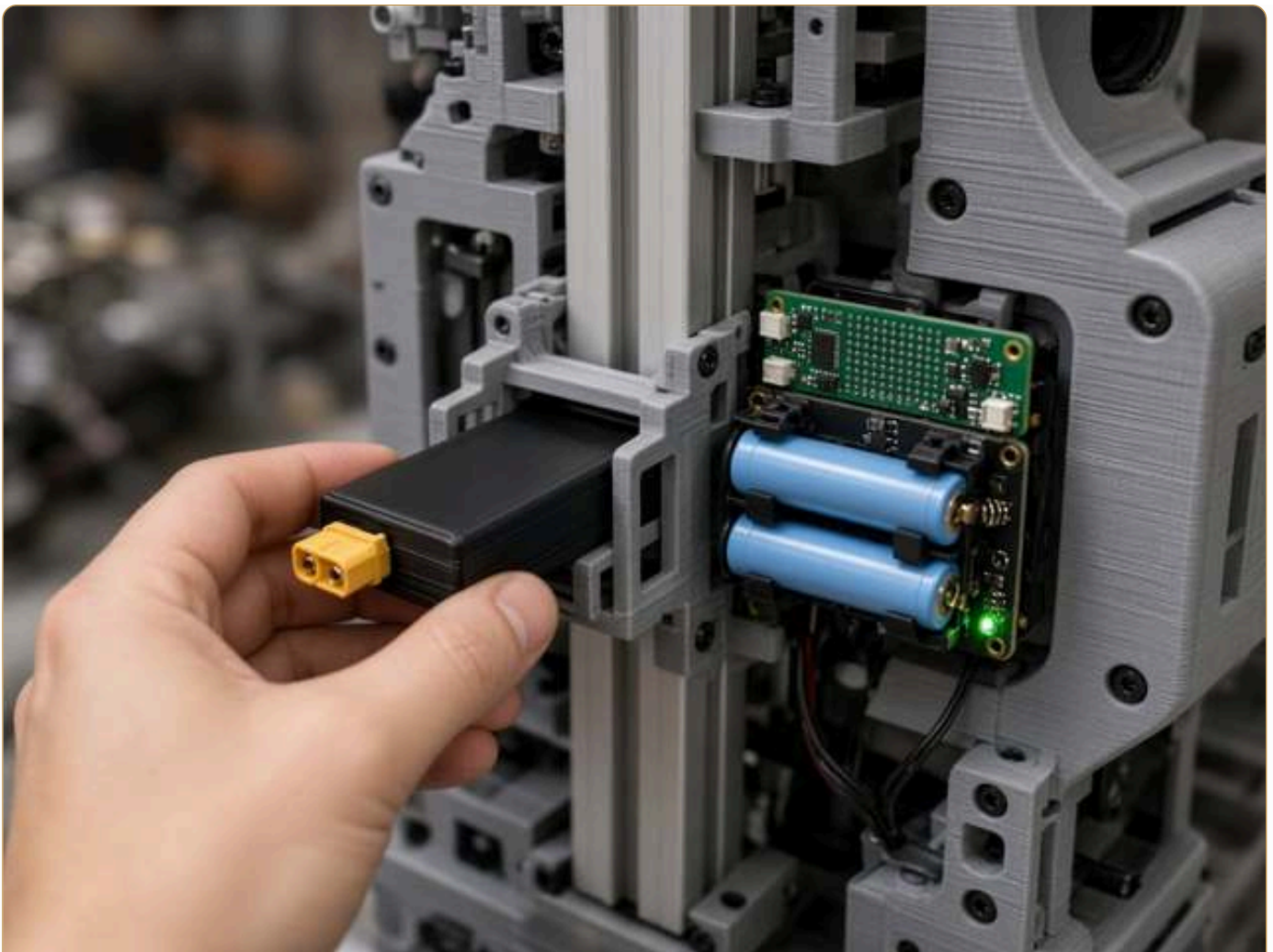
Pull the main battery and Grok keeps talking. A small backup pack keeps the computer, mic, sensors and LTE running for about an hour or more.

Arms park first

It folds its arms to a resting pose and relaxes the motors before power drops, so nothing swings or strains.

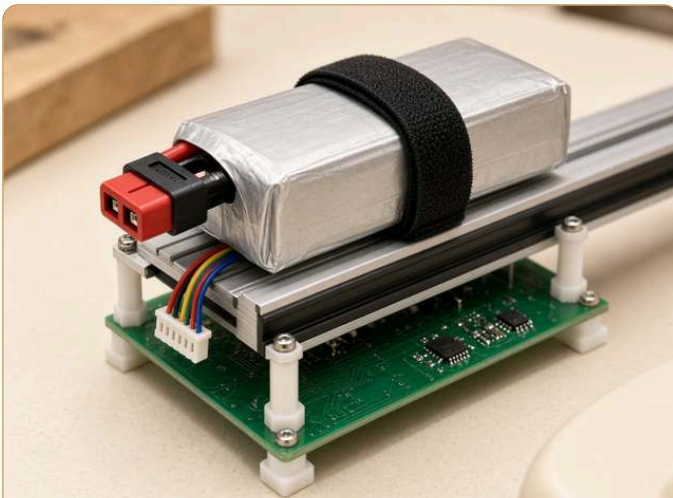
No sparks, clean restart

An anti-spark connector stops the snap on plug-in, and the wheels stay off until the new pack checks out.

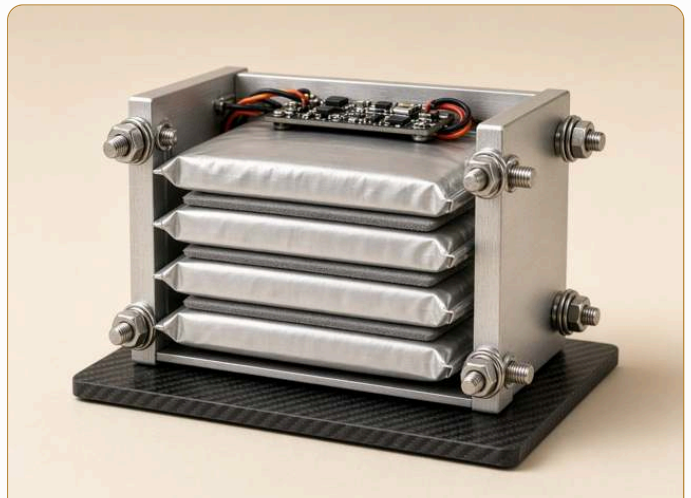


Concept render (AI-generated): swapping the main pack while the X1200-style backup board (two 18650 cells) keeps the brain powered. Generic parts, illustrative only.

Up close: the parts



Concept render: spine pack (flat 3S LiPo shown; the nicer build uses a Bioenno 12.8V LiFePO4 slim pack) on the 2020 spine, with the nylon-standoff thermal break and air gap above the electronics



Concept render: structural option, four LiFePO4 pouch cells (A123 AMP20M1HD-A class) with foam pads, clamped by aluminum end plates, threaded rods and disc springs on a carbon load plate, with a BMS on top



Image: Geekworm. X1200 UPS board under a Raspberry Pi 5 with two 18650 cells (the build specifies Molice P28A cells; the cells pictured are generic)

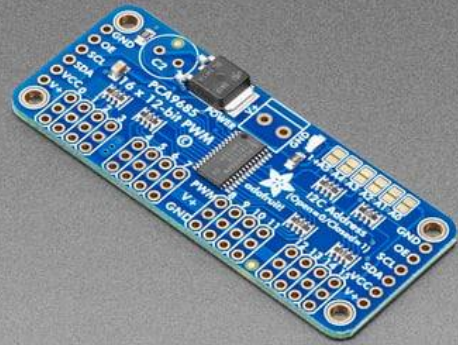
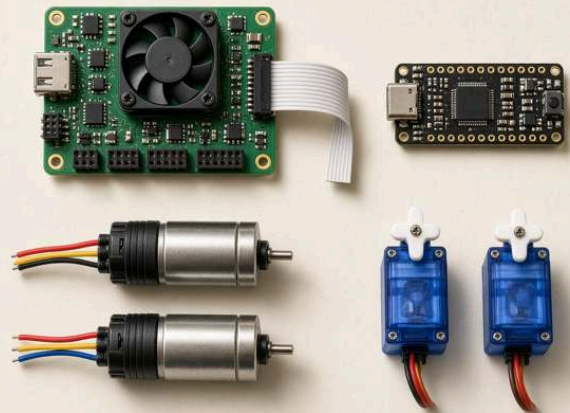


Image: Adafruit. PCA9685 16-channel, 12-bit PWM servo driver (Adafruit 815), which drives the shoulder, gripper and head servos



Concept render: PETG printed motor mount clamping a gearmotor (left) and TPU fin-ray gripper fingers on a servo (right)



Concept render, generic parts (not exact models): an SBC with fan and camera cable standing in for the Pi 5, a dev board standing in for the ESP32-S3, two encoder gearmotors, and two micro servos

What's the difference

	Minimal ~\$615	Nicer ~\$1,265	Structural pack ~\$1,460
Drive	TT/JGB37 gearmotors, TB6612 driver	Pololu 37D encoder motors, 10A Cytron driver	Same as Nicer
Sensing	Pi Camera 3, ultrasonics, MPU-6050	Wide camera, lidar + ToF, BNO085, mic array	Same as Nicer
Power	3S LiPo 5Ah + BMS, FR4 thermal break	Bioenno 12.8V 6Ah LiFePO4, aerogel + fan break	4S A123 LiFePO4, 20Ah (about 260 Wh) in a compression frame

Backup	X1200 UPS keeps the Pi alive	X1200 + ideal-diode OR for the ESP32 and sensors	Same as Nicer
Voice	USB conference mic + 3W speaker; cloud voice, simple wake word; best in a quiet room	XVF3800 4-mic array with on-chip echo cancel; hears across the room and while it is talking	Same as Nicer
Cellular (optional add-on)	+about \$90: USB LTE dongle for SMS alerts and data fallback; video over cellular: snapshots, clips or on-demand live view (no extra hardware)	+about \$120: LTE HAT with two antennas + GPS, SMS, data and voice calls; video over cellular, with better uplink from two antennas	Same as Nicer
Keep-alive	Hot-swap the main pack: brain, voice and LTE stay up about 1-1.5 h on the X1200; arms park and go limp	Same, plus ideal-diode OR on the ESP32 rail and a driver bulk cap	Same as Nicer; optional 2S servo keep-alive (+\$25) holds torque
Weight	Lightest	Moderate	About +1.4 kg; needs a wider base
You gain	The cheapest talking, seeing robot	Mapping, better hearing, a face, safer chemistry	Several times the runtime, and the pack stiffens the spine

▼ **Under the Hood: full specs, parts and build plan**

On this page

Recommendation

Parts list

Torso, shoulders and head

Battery options: structural spine and backup pack

Cost totals

Architecture

Grok API integration

Voice pipeline

Cellular add-on (optional)

Video over cellular (optional)

Keep-alive brain power

Build order

Gotchas

References

Recommendation

Go with two-wheel differential drive plus casters.

Wheels are cheapest, easiest to debug, and work indoors and on patios. Tracks add slip and odometry error; legs multiply servo count, cost, and control complexity. Graduate to tracks only if he wants grass and gravel.

Prices are approximate late-2026 US street prices; Pi and Pololu prices were checked against official pages on 23 Oct 2026, the rest are estimates.

Concept render



Concept render (AI-generated, not a photo): the same full Grok Bot shown in the hero image at the top, at lower resolution. A duplicate, kept here for reference.

Parts list

Minimal viable [Nicer version](#)

Category	Part	Price
Chassis	4WD/2WD acrylic TT-motor chassis kit (Amazon, generic "smart car chassis") or 3D-printed PETG deck	\$20
Motors	2x JGB37-520 12V encoder gearmotor, 110-170 RPM (AliExpress/Amazon)	\$24
Wheels	65-85 mm wheels with 6 mm hubs + 1 ball caster	\$12
Motor driver	TB6612FNG breakout (Adafruit 2448 / SparkFun ROB-14450), 1.2 A/ch	\$6
Real-time MCU	ESP32-S3-DevKitC-1 (Espressif, Digi-Key/Amazon)	\$15
Brain (SBC)	Raspberry Pi 5 2GB (list \$77.50 as of Oct 2026)	\$78
SBC extras	64 GB A2 microSD, Pi 5 Active Cooler	\$17
Camera	Raspberry Pi Camera Module 3 (CSI)	\$25
Range	3x HC-SR04 ultrasonic (use 3.3V-safe HC-SR04P or divider)	\$8
IMU	MPU-6050 breakout (GY-521)	\$5
Mic	USB mini/conference mic	\$15
Speaker	MAX98357A I2S amp (Adafruit 3006) + 3W 4Ω speaker	\$10
Voice extras	Printed chest speaker box + grille, foam/rubber-grommet mic mount in the head, short USB extension (wake word, Vosk/whisper.cpp, Piper are free software)	\$10
Battery	Flat 3S LiPo soft pack, 11.1V 5000-5200 mAh (Ovonic/Zeee class, about 155x48x26 mm), strapped along the back of the spine	\$35
BMS	3S 20A protection BMS between pack and fuse (LiPo packs have none built in)	\$8
Thermal break	1.5 mm FR4/G10 sheet + 10 mm nylon standoffs: an air gap between pack and electronics	\$8
Backup pack (UPS)	Geekworm X1200 Pi 5 UPS (\$39.95-\$48) + 2x Molicel P28A 18650; keeps Pi, ESP32, sensors and Wi-Fi alive	\$60
Keep-alive parts	XT90-S anti-spark connector pair on the main pack lead, plus a 100k/22k VBAT sense divider with a 3.3V TVS into an ESP32 ADC pin	\$8

5V regulation	5V 5A buck for Pi (e.g. Pololu D36V50F5 is better; generic is cheaper)	\$15
Protection	Inline blade fuse 10A, rocker switch, XT60 pair	\$15
Charger	3S balance charger (SkyRC B6 neo class)	\$30
Wiring	16/22 AWG silicone wire, JST-XH, Dupont, heat shrink, standoffs	\$25
Printed parts	PETG filament (motor mounts, sensor brackets, head)	\$20
Cellular (OPTIONAL add-on, not in total)	Waveshare SIM7600G-H 4G DONGLE (\$83.99, antenna included) on USB, plus a 1000 μ F capacitor, short USB cable and a printed clear window	+\$90

3D-print: motor mounts, sensor brackets, battery tray, head/camera pan mount, cable guides. Buy off the shelf: wheels, hubs, bearings, casters, fasteners (M3 assortment), and the battery cells.

Torso, shoulders and head

Budget upper body [Nicer upper body](#)

Category	Part	Price
Torso/neck	2020 extrusion, 4x 300 mm + corner brackets and T-nuts (about 45-55 cm total height)	\$25
Printed shell	PETG torso panels, shoulder brackets, head shell	\$15
Head pan	MG996R servo on a 3-4 in lazy-susan bearing (servo carries no axial load)	\$13
Cable path	12 mm 6-wire capsule slip ring, or a $\pm 170^\circ$ cable loop with soft stops (free)	\$12
Head tilt	MG90S metal-gear micro servo	\$4
Shoulders	2x MG996R (about 9-11 kg·cm at 6V), light printed arms only	\$16
Servo driver	PCA9685 16-ch I2C board (Adafruit 815 or clone) on the ESP32	\$8
Servo power	Separate 6V 10A UBEC/buck (e.g. Hobbywing 10A UBEC class) + 1000 μ F cap	\$15
Spine battery mount	Printed PETG cradle on the rear 2020 rails + 2 rubberized straps	\$6
Extras	Servo extensions, metal horns, M3/M5 hardware	\$10

Budget upper body adds about \$124, nicer adds about \$216; all servo prices are approximate.

Battery location: flat pack on the back of the spine, low and centered.

The pack stands upright against the rear 2020 rails, its bottom edge 2-3 cm above the deck and centered left-to-right on the robot's centerline, so its mass sits over the axle line. Controller boards (Pi, ESP32, buck, servo BEC, fuse block) mount on the front face of the spine, with the thermal break between them and the pack: standoffs, an FR4/G10 sheet (plus an aerogel strip on the nicer build), and a vertical air channel open at the bottom and top.

	LiPo soft pack (budget)	LiFePO4 pack (nicer)
Voltage	3S: 11.1V nominal, 12.6V full, cut off at 9.9V	4S: 12.8V nominal, 14.6V full, cut off at 10-11V; flat curve suits 12V motors
Energy per kg	Higher (about 150-200 Wh/kg)	Lower (about 90-120 Wh/kg), heavier for the same runtime
Safety	Puncture or overcharge can mean thermal runaway and fire; needs an added BMS and a LiPo charging bag	Much more thermally stable, far less fire risk; BMS built in

Cycle life	About 300-500 cycles	About 2,000+ cycles
Charger	LiPo balance mode	LiFe mode (3.65V/cell); do not charge on LiPo settings

- Center of gravity: keep the CG below about a third of total height and on the centerline. The spine pack weighs about 0.35 kg (LiPo) to 0.9 kg (LiFePO4); mount it as low as the spine allows and centered, so when both arms swing forward the CG still stays inside the wheel footprint. Trim with ballast on the front of the deck if the robot leans back.
- Thermal: the pack runs warm under servo load and the Pi 5 and bucks run hot. Keep at least a 10 mm gap, vent the channel bottom-to-top, and log pack temperature; stop charging or driving above 45°C (LiPo) or 55°C (LiFePO4).
- Wheelbase: grow track width to at least 60% of robot height (about 28-30 cm for a 50 cm robot), and add casters front and rear.
- Acceleration: cap at about 0.3 m/s² in ESP32 firmware or the torso tips forward on stops.
- Battery: four MG996R can stall at 2.5A each; budget 3-5A extra and a dedicated servo rail so the Pi never browns out.
- Head: add tools `look_at(pan, tilt)` and `nod()`; shoulders get `wave()` and `point(direction)` with clamped angles.

Battery options: structural spine and backup pack

⚠ True structural batteries are still lab hardware.

Cells that carry load themselves, like Chalmers University's carbon-fiber structural battery, are research prototypes with very low energy density, and you can't buy them. The DIY version is a battery-as-structure module: off-the-shelf LiFePO4 pouch cells clamped in a compression frame, where the frame and a sandwich plate carry all the load and the cells never see bending.

Part	Spec / source	Price
Cells	4x A123 AMP20M1HD-A LiFePO4 pouch, 3.3V 20Ah, 65Wh, 7.25 x 160 x 227 mm, about 496 g each; 4S gives 13.2V, 260 Wh, about 2 kg. Authorized-reseller listing about \$45/cell; Alibaba \$12-25 (authenticity varies)	\$180
Alternatives	EVE LF50K or CALB prismatic LiFePO4 (aluminum can, also wants compression); heavier and thicker, best for a bigger base	approx.
End plates	2x 6061 aluminum, 6 mm (3/16-1/4 in), slightly larger than the cell face	\$20
Tension	4x M6 stainless threaded rod + Belleville disc-spring stacks (or 2 steel banding straps)	\$12
Load plate	2-3 mm carbon fiber or 3 mm aluminum sandwich plate bolted to the 2020 spine; carries all bending and arm loads	\$35
Compliance	PORON/silicone foam pads between cells (absorbs swelling, spreads pressure)	\$10
BMS	4S LiFePO4 smart BMS, 40-60A (JBD/Daly class), with temp probe	\$25
Fusing	ANL or MIDI fuse at the pack + per-branch blade fuses	\$5

- Compression: A123's pack design guide recommends 4-18 psi evenly on the broad faces, with 8-14 psi used for its 3,000-cycle rating. The 160 x 227 mm face is about 56 in², so 10 psi is about 560 lbf total, or about 140 lbf per M6 rod. Disc springs keep the pressure steady as cells grow.
- Swelling: cells grow about 1% in thickness each cycle and 3-5% over their life. Leave that travel in the springs and foam, never in rigid steel.
- No puncture or flex: cover the stack's edges and tabs, keep every screw out of the cell zone, and never let the cells act as a beam. If it bends, the plate is too thin.
- Weight: about 2.3 kg with the frame, versus about 0.9 kg for the Bioenno pack. Only choose this if you want long runtime and can widen the base; it pushes the robot past 4 kg.
- Cost: about \$287 in place of the Bioenno pack and cradle (\$88), so the nicer build goes from about \$1,265 to about \$1,460.

Auxiliary backup pack

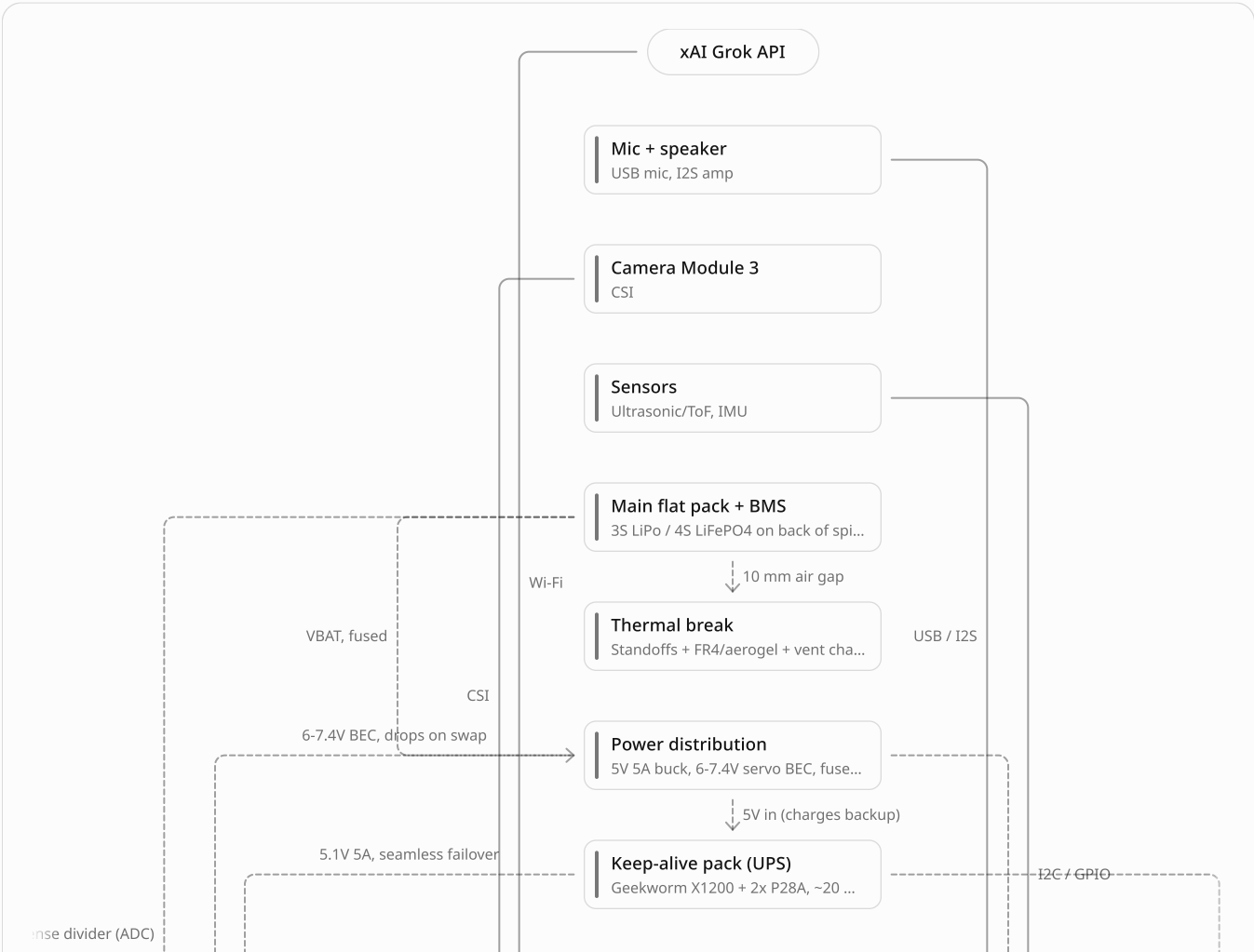
Part	Geekworm X1200 Pi 5 UPS (\$39.95 at Central Computer, \$48 at Geekworm) + 2x MoliceL P28A 18650 (about \$6 each). It outputs 5.1V 5A, sits under the Pi, and uses pogo pins, not the 40-pin header.
Failover	The X1200 takes 5V from the main buck, charges the cells at up to about 3A, and switches to battery automatically when input drops. The ESP32, sensors, and USB mic draw from the Pi's 5V rail, so they stay up too.
Nicer: ideal-diode OR	A TI LM66100 ideal-diode breakout (or LTC4412-based module) ORs the main 5V buck and the UPS 5V into the ESP32/sensor rail, so no rail ever back-feeds the other
Shutdown	The X1200 flags power loss on a GPIO and reports cell voltage over I2C. A systemd service tells Grok "main power lost", stops the motors, then runs shutdown -h at about 3.3V/cell or after a backup timer (now 45 minutes by default, so a planned swap has room; the earlier 10-minute timer was too short for hot-swap)
Swap	With the backup in place, the main pack can be swapped without rebooting the Pi or dropping the Grok session
Alternative	Waveshare UPS HAT (B or E) for other Pi boards; check that it can supply 5A before using it with a Pi 5

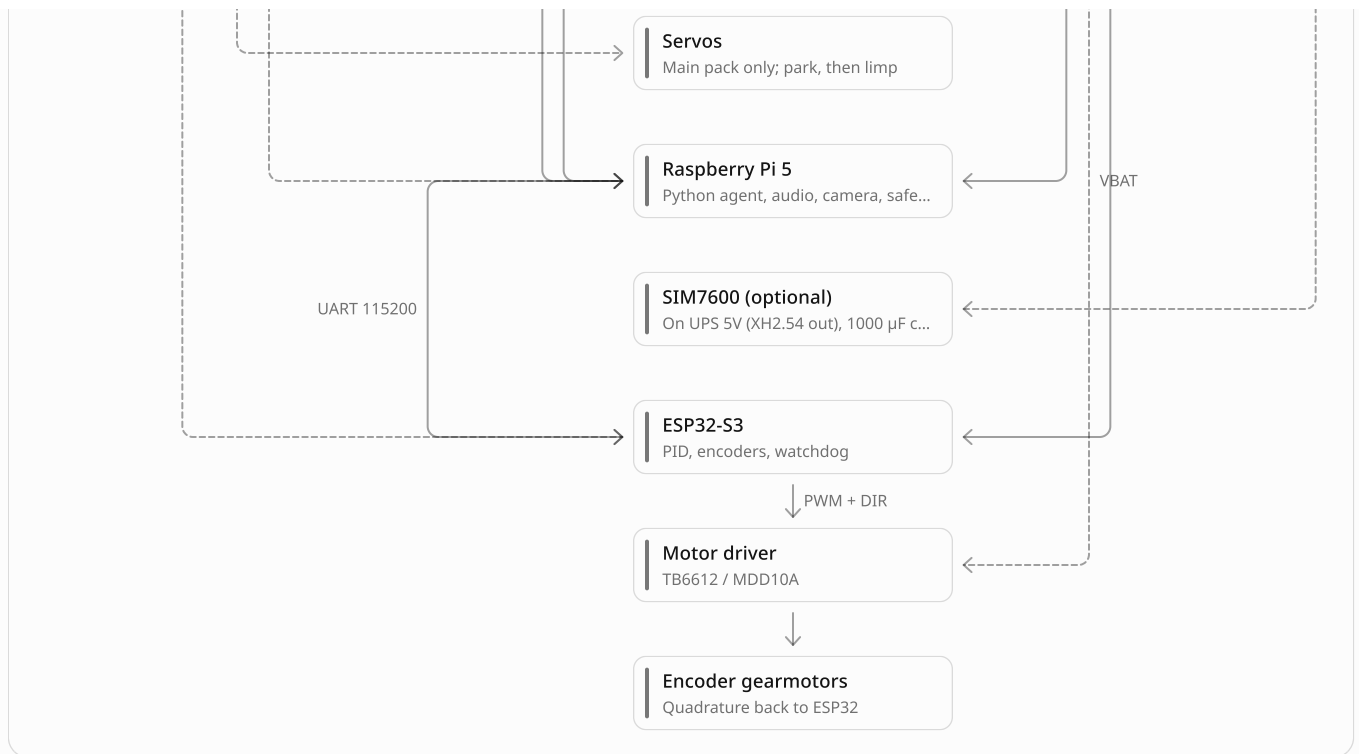
Cost totals

Build	Hardware	Running cost
Minimal viable	\$570 to \$660 (about \$615)	Voice about \$0.08/min
Nicer version	\$1,165 to \$1,365 (about \$1,265)	Voice about \$0.08/min plus vision calls
Nicer + structural spine pack (option)	about \$1,460	Swaps the Bioenno pack and cradle for a 4S A123 compression-frame pack

Includes the torso and head. Biggest swings: Pi RAM tier, Pololu vs AliExpress motors, lidar, and servo choice.

Architecture





- Pi 5 talks to the cloud and does audio and camera; it never drives motors directly.
- ESP32 runs the 100 Hz loop: wheel PID, encoder odometry, range sensors, and a stop if no command for 300 ms.
- UART protocol: simple line JSON, e.g. `{"v":0.2,"w":0.0,"ttl":500}` down and telemetry up at 10-20 Hz.
- Common ground between Pi, ESP32, and driver; motor power runs on its own thick wires from the fuse.

Grok API integration

Voice in/out	Speech-to-speech over WebSocket <code>wss://api.x.ai/v1/realtime?model=grok-voice-latest</code> , server VAD, PCM audio, about \$0.08/min
Auth	API key in <code>XAI_API_KEY</code> on the Pi only (server-side); never in firmware
Tools	Declare functions in <code>session.update</code> : <code>drive(v,w,seconds)</code> , <code>turn(degrees)</code> , <code>stop()</code> , <code>look()</code> , <code>get_status()</code>
Vision	<code>look()</code> grabs a JPEG (640x480 is plenty), sends it to a vision-capable Grok model via the REST chat API, returns a text description to the voice session
Safety gate	Pi clamps speed, rejects moves when ToF/ultrasonic under 25 cm, caps duration; ESP32 halts on UART silence; hardware E-stop cuts motor power

- 1 Listen**
Mic audio streams to the realtime session; Grok replies with audio deltas played on the speaker.
- 2 Decide**
When the user says "come here" or "what do you see", Grok emits a function call.
- 3 Act**
Pi validates the call, sends a time-limited command to the ESP32, and returns the result (or sensor refusal) as the function output.
- 4 Repeat**
Push short telemetry text (battery, obstacle distances) into the session so Grok knows its state.

Check docs.x.ai for current model names before coding; the voice alias was `grok-voice-latest` (Grok Voice 3.0) as of Oct 2026.

Voice pipeline

Mic (minimal)	USB conference/mini mic, about \$15, already in the BOM. Optional upgrade: Sreed ReSpeaker 4-Mic Array for Raspberry Pi HAT (about \$27-35, no on-board echo cancel).
Mic (nicer)	Sreed reSpeaker XVF3800 USB 4-Mic Array, \$60.99 bare / \$62.99 with case. Alternative: ReSpeaker USB Mic Array v2.0 (XVF3000), \$64. Both do beamforming, AEC, noise suppression and DoA on-chip and show up as a USB sound card.

Speaker	Minimal: MAX98357A I2S amp + 3W 4Ω speaker (I2S on GPIO 18/19/21). Nicer: class-D amp + 5W speaker, or a USB speaker. Mount it in a sealed printed box in the chest.
Wake word	openWakeWord (free, Apache-2.0, custom words trainable) or Picovoice Porcupine (free personal tier, needs an access key), on the Pi 5 CPU at a few percent load.
Primary path	After the wake word, stream mic PCM to the Grok realtime voice WebSocket (server VAD) and play the audio deltas back. Typically about 0.5-1.5 s to first audio on good Wi-Fi; about \$0.08/min.
Offline fallback	STT: Vosk (streams in real time, less accurate) or whisper.cpp tiny/base (about 1-3 s per short sentence on a Pi 5). LLM: llama.cpp or Ollama with Llama 3.2 1B/3B or Qwen2.5 1.5B/3B at Q4; roughly 8-12 tokens/s for 1B and 3-5 tokens/s for 3B, so 5-15 s per short answer. TTS: Piper, under 1 s per sentence. Expect 5-20 s per turn: it works, but it is not conversational. The 2GB Pi in the minimal build should stick to 1B models; the 4GB Pi can run 3B.

- Placement: mics on top of the head, as far from the drive motors and the chest speaker as possible; soft-mount the board on foam or grommets so gearbox vibration does not reach it.
- Echo cancellation: the XVF3800/v2.0 need the playback signal as a reference. Route speaker audio through the array's own output (3.5 mm/USB playback) or feed its reference input, otherwise AEC cannot work. With the cheap USB mic, mute the mic while the robot talks (half duplex).
- Barge-in: with working AEC, leave the mic open while it talks; when server VAD detects speech, stop playback immediately and cancel the current response. Without AEC, use the wake word or a button to interrupt.
- Motor and servo noise: gearmotors and servos are loud to a mic a few cm away and confuse VAD. Pause or slow motion while listening, add ferrites and a separate 5V rail for the servos, and twist the motor leads. PWM whine can leak into an analog amp; I2S/USB audio avoids that.
- Power: put the amp on the 5V buck, not the servo rail; volume peaks can brown out the Pi if they share a weak supply.
- Privacy: nothing is streamed to the cloud until the local wake word fires; show a light on the mic array while it is listening.

Cellular add-on (optional)

Nicer pick	Waveshare SIM7600G-H 4G HAT, \$97.99 (HAT (B) \$110.99, M.2 HAT \$112.99) as of Oct 2026. Global LTE Cat-4 (150/50 Mbps), 3G/2G fallback, GNSS, SMS, and voice through its audio jack or USB audio. Connect it over USB to the Pi 5; GPIO pass-through is fine.
Minimal pick	Waveshare SIM7600G-H 4G DONGLE, \$83.99 on Waveshare's product page (another listing shows \$100.99), antenna included. Plugs into USB; same SIM7600G-H module and SMS; voice calls are possible over AT commands but awkward, so treat it as SMS + data.
Alternative	Quectel EG25-G mini-PCIe, \$79.99 (module only), plus a mini-PCIe-to-USB adapter with SIM slot (about \$15-25) and antennas (about \$15): about \$110-120, LTE Cat-4 global with US bands 2/4/5/12/13/25/26, SMS and voice over USB audio.
Cat-1 note	Waveshare's SIM7600E LTE Cat-1 HAT (\$84.29-85.99) is the Europe/Asia/Africa version, so it is not a good US choice and is barely cheaper. For the US, stick with SIM7600G-H or EG25-G; Cat-1 is plenty fast for SMS and voice data, but not cheaper in practice here.
US carriers	SIM7600G-H and EG25-G work on AT&T and T-Mobile (and MVNOs on them, like Mint and Tello) on LTE. Verizon often rejects modules whose IMEI is not on its approved list, so avoid Verizon-based plans. 2G/3G are shut down in the US, so it must stay on LTE (VoLTE for calls; check that the firmware and carrier support VoLTE on the module).
SIM	Nano SIM in the HAT's slot (the dongle has its own slot). Leave a printed access door so the SIM can be swapped without opening the torso.
Antennas	Two LTE antennas (main + diversity/aux) plus the GNSS patch on the G-H. Mount them high, in the head or across the back of the shoulders, at least 5-10 cm from the drive motors and off the aluminum spine (metal detunes and blocks them). Put them behind a clear PETG/polycarbonate window, not behind metal or carbon; the GNSS patch faces the sky.
Power	LTE transmit peaks reach about 2 A at the module. Feed it from the 5V rail (Pololu buck), not the Pi's USB alone, with a 1000 μF low-ESR capacitor near the board. Put it on the X1200-backed rail so it stays alive and can send the alert when the main pack trips.

Plan	Type	Approx. monthly	Best for
Hologram	IoT data SIM (SMS via API)	from \$1.50/device + usage; alerts-only about \$2-5	Alerts and telemetry
Soracom	IoT data SIM	pay-as-you-go from about \$0.002/MB; alerts-only about \$1-5	Alerts and telemetry
Twilio Super SIM	IoT data SIM	a few dollars per SIM + per-MB; check current rates	Alerts with Twilio SMS/APIs
Tello	Prepaid voice/SMS/data (T-Mobile)	\$5 custom; \$10 for 2 GB + talk and text	SMS to Max and voice calls
Mint Mobile	Prepaid voice/SMS/data (T-Mobile)	\$15/mo intro or with 12-month prepay; then \$25+	More data for voice fallback

- Alerts: the Pi sends a text to Max for a safety-gate trip or E-stop, a tip or fall (BNO085/MPU-6050 tilt over about 45°), low battery, or stuck (motors on, no encoder or position change for 10 s). Rate-limit to one text per event type every few minutes.
- Data fallback: when Wi-Fi drops, NetworkManager/ModemManager switches the default route to the LTE interface (wwan0) and the Grok voice session reconnects. Realtime voice uses roughly 3-6 MB per minute, so 30 minutes a month is about 100-200 MB.
- Remote calls: with a voice plan, Max can call the robot's number; the Pi answers and bridges the call audio to the speaker and mic (or to the Grok session). This needs the HAT or EG25-G with VoLTE working on the carrier; test it before relying on it.
- Privacy: a cellular robot can listen and report from anywhere. Keep the wake word local, show an LED when the modem or mic is active, allow inbound calls only from Max's number, and do not stream video over LTE by default.
- Cost: cap data in the plan or with a usage alert; a stuck voice session over LTE can burn data quickly. An IoT SIM for alerts only is the cheapest choice; add a prepaid voice plan only if remote calls matter.

Video over cellular (optional)

Can it stream?	Yes. The SIM7600G-H is LTE Cat-4: 150 Mbps down / 50 Mbps up in theory. Real US uplink is typically about 2-20 Mbps and depends on signal, so a 1-2 Mbps stream fits with headroom; plan for drops to under 1 Mbps at the edge of coverage.
Interface	USB 2.0 (480 Mbps) to the Pi 5, so USB is never the bottleneck. It shows up as a network interface via RNDIS/ECM (usb0) or QMI (wwan0, with ModemManager/NetworkManager or qmicli). No extra hardware.
Encoding	The Pi 5 has no hardware H.264 encoder; rpicas-vid uses libx264 in software. Raspberry Pi's own white paper measures 1080p30 at about 60% of one core (low-latency mode) to 100-150% of one core (high quality). 640x480-720p at 15 fps and 1-2 Mbps uses well under one core, leaving room for voice and the robot loop.
(a) Snapshots + clips	On a motion or alert event, save a JPEG or a 10-30 s H.264 clip and upload it to Google Drive (rclone), S3 or Backblaze B2; Max opens the link from the SMS. Cheapest on data and battery; delay is the upload time (a few seconds).
(b) Live WebRTC	MediaMTX or go2rtc on the Pi serves WebRTC. Carrier CGNAT blocks inbound connections, so put the Pi and Max's phone on Tailscale (free Personal plan: up to 6 users, unlimited devices), or use a TURN/relay service. WebRTC latency is typically about 200-500 ms; HLS or RTSP through a relay is about 2-10 s.
(c) On-demand only	Recommended default: the stream starts only when Max asks (via a Tailscale web page or a text command), auto-stops after 5 minutes, and an LED shows it is live. Saves data, power and privacy.

What	Size / rate	Data	Plan impact
Snapshot (640x480-720p JPEG)	about 50-200 KB	100 snapshots = 5-20 MB	Fine on any plan, even IoT
30 s clip (720p, 1-2 Mbps)	about 3-8 MB	20 clips = 60-160 MB	Fine on a 2 GB plan
Live, 1 Mbps	about 450 MB/hour	10 h/month = 4.5 GB	Needs a 10 GB plan
Live, 2 Mbps	about 900 MB/hour	10 h/month = 9 GB	10-20 GB plan

Plan	Price (Oct 2026)	Fit for video
Tello 2 GB + talk/text	\$10/mo	Snapshots and clips
Tello 10 GB	\$15/mo	A few hours of live view
Tello 20 GB	\$20/mo	Regular live view
Tello Unlimited	\$25/mo (slower after 50 GB)	Heavy use
Mint Mobile (any tier)	\$15/mo intro or 12-month prepay (\$180); regular 3-month rates \$25-40	Good with annual prepay
IoT SIM (Hologram/Soracom)	per-MB pricing	Alerts and snapshots only; live video gets expensive fast

- Power: the modem uses about 0.5-1 W idle and about 2-3 W while uploading; encoding adds roughly 1-2 W on the Pi. Streaming adds about 3-5 W total, roughly 3-5 Wh per hour, or about 5-7% of a 55-77 Wh main pack per hour of live view.
- Added cost: \$0 hardware; MediaMTX, go2rtc, rclone and Tailscale Personal are free. Cloud storage is pennies at this scale (Backblaze B2 or S3 at a few GB) or free within a Google Drive quota. The real cost is the data plan: from about \$10/mo for clips to \$15-25/mo for regular live view.

- Latency for driving: 200-500 ms WebRTC is fine for looking around and slow driving with the safety gate on; never drive remotely at speed over HLS (2-10 s).
- Privacy: on-demand only by default, Tailscale-only access (no public port), LED on while streaming, clips auto-delete after 7-30 days, and no recording inside bedrooms or bathrooms.

Keep-alive brain power

Standard on every build. The X1200 already in the BOM becomes the formal keep-alive pack: it carries the brain through a main-pack swap, while anything that moves is allowed to drop.

X1200 (verified, Geekworm wiki)	5.1V ±5%, 5A max output to the Pi 5 through pogo pins. Two 18650 slots in parallel (1S), each with its own 5A fuse. Charge current 2.3-3.2A max, terminal 4.23V, recharge at 4.1V. Seamless switching between input and battery. Power loss flagged on a GPIO (GPIO6 in Geekworm's X120X scripts); MAX17040 fuel gauge at I2C 0x36. Two extra 5V outputs on XH2.54 2-pin connectors.
Pack energy	2x Molice! P28A (2,800 mAh, 3.6V) = about 20 Wh. Use 80% and about 90% boost efficiency: about 14.5 Wh usable.
What stays up	Pi 5 (5-8 W), mic and speaker idle (about 1 W), SIM7600 (1-3 W), ESP32 + sensors (about 1 W): 8-12 W, so about 1.2-1.8 h. Plan on an hour or more.
Feeding peripherals	USB mic/camera: from the Pi's USB ports (1.6A total with PSU_MAX_CURRENT=5000 set in EEPROM). Modem: from an X1200 XH2.54 5V output with its 1000 µF cap. Everything shares the 5A budget; worst case about 3.5-4A with LTE peaks.
Not on the X1200	Servos and drive motors. An MG996R or DS3218 stalls at about 2.5-3A each, so two shoulders alone could exceed the 5A rail and brown out the Pi.
Shutdown thresholds	Unchanged floor: shut down at about 3.3V/cell. The backup timer moves from 10 to 45 minutes (configurable) so a swap is never rushed.

Event	What happens	Hardware
Planned swap	Say “swap battery” or press the swap button: arms move to the park pose, PCA9685/bus servos go limp, wheels stop, then pull the pack	Park pose rests on printed arm stops
Unplanned loss	Pi sees the X1200 power-loss GPIO; ESP32 sees VBAT fall below about 9V (3S) or 10V (4S) and safe-stops the drivers within one control loop	100k/22k divider + 3.3V TVS on an ESP32 ADC
Brownout	Pi: seamless X1200 handoff. Modem: rides the UPS rail plus 1000 µF. ESP32 (Nicer): LM66100 ideal diode picks the UPS 5V. Motor drivers lose power	Existing LM66100, modem cap
Reconnect	XT90-S resistor contact precharges the driver and BEC caps, so no spark. ESP32 waits for VBAT stable for 1 s, re-enables the driver (TB6612 STBY / MDD10A PWM at zero), re-homes servos slowly	XT90-S anti-spark, 470 µF driver cap (Nicer)

Optional: servo keep-alive (+\$25, not in totals).

A small 2S 1,000 mAh LiPo through an ideal diode into the servo BEC input lets the arms hold low torque through a swap. Only worth it if the arms carry something; parking and going limp is the default.

Build order

- 1

1. Bench power
Build the pack, BMS, fuse, switch, buck. Verify 5.1V under a 4A dummy load before any board sees it.
- 2

2. Pi alone
Flash Pi OS, cooler on, camera test with rpikam-hello, mic/speaker test.
- 3

3. Voice first
Python script to the xAI realtime endpoint on the bench; talk to it before it has wheels.
- 4

4. ESP32 + one motor

Solder driver headers, spin one motor, read encoder counts, tune PID on the stand.

5

5. Chassis

Print mounts, fit both motors and casters. The main battery no longer sits on the deck; it moves to the spine in step 10.

6

6. UART link + watchdog

Pi to ESP32 commands; pull the cable and confirm the motors stop within 300 ms.

7

7. Sensors

Add ultrasonic/ToF and IMU on the ESP32; obstacle stop runs locally.

8

8. Tools

Wire Grok function calls to drive/turn/stop/look; test wheels-up first.

9

9. Servo rail

Wire the 6V/7.4V BEC with its own fuse, PCA9685 or bus adapter to the ESP32; center every servo before fitting horns.

10

10. Torso, neck and spine battery

Bolt the 2020 frame to a widened base. Battery location: mount the flat pack cradle on the rear rails, bottom edge 2-3 cm above the deck, centered on the centerline. Fit standoffs and the FR4 (plus aerogel on the nicer build) barrier; leave the 10 mm vent channel open top and bottom, and put the Pi, bucks and fuse block on the front face. Push-test for tipping with arms forward before power.

11

10b. Backup pack

Fit the X1200 under the Pi 5 with 2x 18650, feed it 5V from the main buck, and power the ESP32 and sensors from the Pi's 5V header. Pull the main XT90-S: the Pi must stay up, and a low-battery event must trigger a clean shutdown.

12

11. Pan/tilt head

Bearing, pan servo, slip ring or cable loop, tilt servo, camera; set soft angle limits in firmware.

13

12. Shoulders

Mount shoulder servos, light arms, test gestures at low speed.

14

13. Shell and polish

Printed panels, face display, cable management, outdoor sealing if needed.

Gotchas

Thermal

- Pi 5 throttles at 85°C without the Active Cooler; give it an air path.
- TB6612 is 1.2A continuous; stalls on carpet exceed it. Size driver to motor stall current.
- PLA softens in a hot Louisiana car or patio; print in PETG or ASA.

Battery safety

- Soft LiPo packs have no can: protect the spine pack with a printed cover so nothing can puncture it, and never let it flex. LiFePO4 is the safer choice in a home.
- Charge the LiPo in a LiPo bag off the robot; charge LiFePO4 only in LiFe mode.
- Inspect for swelling monthly; a puffed pouch gets retired.
- Fuse right at the pack; balance-charge on a non-flammable surface, never unattended.
- Buy cells from 18650battery store-type vendors; fake high-mAh cells are common on marketplaces.
- Low-voltage cutoff at 3.3V/cell in firmware too.

Weight vs torque

- Target under 4 kg total; budget motors are fine to about 3 kg.
- Required torque per wheel is roughly mass x g x wheel radius x 0.3 for ramps and carpet.
- Keep battery low; a tall head on a short wheelbase tips on stops.

Humidity and outdoor

- Shreveport humidity condenses inside boxes overnight: add silica gel and a vent.
- Conformal-coat the ESP32 and driver; leave connectors and the camera lens clean.
- Ultrasonic sensors misread in rain and wet grass; ToF struggles in bright sun.

- Pi 5 can pull 5A peaks; a weak buck causes random brownout reboots.

- Treat it as splash-resistant, not waterproof; no puddles.

References

	xAI Voice overview (pricing, realtime) docs.x.ai	
	xAI Speech to Speech (realtime WebSocket) docs.x.ai	
	Pololu 50:1 37D 12V gearmotor with encoder pololu.com	
	Pololu Dual TB9051FTG driver for Pi (alt. driver, \$39.95) pololu.com	
	Raspberry Pi Oct 2026 price change raspberrypi.com	