Telemedicine Video Scaling & Surge Capacity Checklist

Pressure-test your video capacity plan before a seasonal surge does. Engineering guidance, not legal advice — confirm specifics with counsel.

1 · CAPACITY MATH (know your busy minute)

- [ ] Estimate concurrency = daily visits × avg minutes ÷ clinic minutes (not total/day)
- [ ] Outbound per 2-person consult ≈ 2 × 1.5 Mbps = 3 Mbps from the media server
- [ ] One SFU tops out ~1,000-1,600 concurrent consults at ~50% pipe use — plan past it
- [ ] Apply a 2-3x midday peak factor; add a further margin for a flu/PHE surge
- [ ] Right-size video quality to the clinical task — don't ship 1080p where 480p is diagnostic

2 · HORIZONTAL SCALING (many servers, one router)

- [ ] A session router assigns each new consult to an SFU with spare capacity
- [ ] Routine consults each live on ONE SFU — scale by adding servers
- [ ] Cascade SFUs only for one very large room (grand rounds, all-hands), not visits
- [ ] Load balancer + shared session/state store across the SFU pool
- [ ] Never run an SFU near 95% — a single failure must not cascade

3 · REGIONS & DATA RESIDENCY (latency + law)

- [ ] Route each patient to the NEAREST media server (target < ~150 ms one-way, ITU-T G.114)
- [ ] HIPAA does NOT require PHI to stay in the US — safeguards + BAA + breach rules govern
- [ ] Foreign law often DOES restrict it: GDPR special-category data, national residency rules
- [ ] Draw the region map and the legal map together, before onboarding a new country
- [ ] Encrypt every hop that crosses a system boundary in transit

4 · SURGE & AUTOSCALING (ready for the worst Monday)

- [ ] Provision ABOVE the surge peak, not the daily average
- [ ] Predictive autoscale: scale up ahead of the morning rush and the flu curve
- [ ] Keep WARM headroom — a new media server is useless until patients route to it
- [ ] Multi-region failover: shift a region's traffic elsewhere if it goes dark
- [ ] Degrade gracefully: drop resolution, then audio-only — never a dropped consult

AVAILABILITY IS A HIPAA GOAL (not just an SLA)

The HIPAA Security Rule protects the confidentiality, integrity, AND availability of electronic PHI (45 CFR 164.306(a)) — availability has equal standing with privacy. A platform that collapses under a surge, or loses a region with no failover, strains the availability goal. Two requirements make it operational: the contingency-plan standard (45 CFR 164.308(a)(7)) requires a data backup plan, a disaster recovery plan, and an emergency mode operation plan; and technical safeguards (45 CFR 164.312(a)(2)(ii)) require an emergency access procedure so authorized clinicians can still reach needed ePHI. So multi-region redundancy and surge headroom are part of your contingency posture, not only your uptime number. Note: the 2026 HIPAA Security Rule update (NPRM 90 FR 800, RIN 0945-AA22) would tighten contingency and resilience controls — proposed, not final as of mid-2026; confirm status before relying on a deadline.