Compressed Feasibility: A Technical Update to the Ambient Non-Consensual Synthesis Threat Model

Travis Gilly
Real Safety AI Foundation
t.gilly@ai-literacy-labs.org

Abstract

In January 2026, we published a convergence threat analysis demonstrating that ambient non-consensual intimate image synthesis via AR wearables was architecturally feasible through cloud-assisted pipelines and would achieve edge-only feasibility within 12 to 24 months [1]. This technical update reports that multiple independent developments in the eight weeks following publication have compressed our timeline estimates significantly. Specifically, we identify three capabilities released between February and March 2026 that collectively reduce barriers across the critical pipeline stages: person segmentation (MatAnyone2, 140MB model achieving state-of-the-art video matting), inference acceleration (Diagonal Distillation, achieving 270x speedup over baseline video generation), and mobile 3D rendering (MobileGS, achieving 120+ FPS Gaussian splatting on a Snapdragon 8 Gen 3 phone at 4.8MB model size). We revise our feasibility classes accordingly, noting that the "edge full video synthesis" class we originally projected at 12 months may now be achievable within 6 to 9 months under moderate assumptions. We discuss implications for the policy recommendations in our original analysis and argue that the structural mismatch between capability maturation and regulatory response has widened.

1. Introduction

Our original threat analysis [1] established four feasibility classes for ambient non-consensual intimate image synthesis via AR wearables:

1. Cloud-assisted real-time (feasible at time of publication) 2. Edge static image synthesis with interpolation (feasible at time of publication) 3. Edge full video synthesis (projected at 12 months under moderate assumptions) 4. Edge real-time continuous synthesis below 500ms (projected at 24 to 36 months)

We noted that the direction of development was clear and that our directional conclusions did not depend on precise timeline accuracy [1, Section 3.5]. We also identified key dependencies: one-step diffusion optimization, mobile NPU advancement, model quantization and distillation research, and world model architectures.

In the eight weeks since publication, independent research groups have released open-source tools that advance each of these dependencies. None of these tools were developed for malicious purposes. Each addresses a legitimate technical challenge. Their significance for our threat model lies not in individual capabilities but in what they collectively enable when composed into a pipeline.

This update does not repeat the threat model, legal analysis, or philosophical grounding of the original paper; readers unfamiliar with the ambient synthesis threat should consult [1] for full context. Here we

focus narrowly on what has changed technically and what those changes mean for feasibility timelines.

2. New Capability Developments

2.1 Person Segmentation: MatAnyone2

MatAnyone2 [2] is an open-source video matting model that achieves state-of-the-art performance in separating people from video backgrounds, including under challenging conditions such as rapid motion, complex hair, and occluding objects. The model produces clean alpha-channel masks suitable for downstream compositing pipelines.

Three properties are relevant to our threat model:

First, the model is extremely small. At approximately 140 megabytes, it fits comfortably on mobile and wearable-class hardware. For comparison, full diffusion models typically require 2 to 8 gigabytes of storage and proportional memory during inference. A 140MB segmentation model running as a pre-processing stage adds negligible overhead to the overall pipeline.

Second, performance quality exceeds prior tools. Compared to existing matting solutions such as GVM, MatAnyone2 produces significantly higher resolution masks with cleaner edge handling, particularly around hair and fine features. This is directly relevant because the original pipeline Stage 3 (region crop and pose estimation) depends on accurate person segmentation for the synthesis stage to produce convincing output. Poor segmentation produces artifacts at body boundaries; high-quality segmentation enables seamless compositing.

Third, the tool is fully deployed. Open-source code is available on GitHub with installation instructions, and a free HuggingFace Space provides browser-based access. No technical barrier to acquisition exists.

In our original pipeline latency table [1, Section 3.2], we estimated person detection, tracking, and region cropping at 35 to 60ms combined. MatAnyone2 operates in the same latency range while producing substantially higher quality segmentation output. This does not change the timing of Stage 3 but improves the quality of input to Stage 4 (synthesis), reducing the quality gap between cloud-assisted and edge-only configurations.

2.2 Inference Acceleration: Diagonal Distillation

Diagonal Distillation (DiagDistill) [3] is a novel distillation technique for video diffusion models that achieves a reported 270x speedup over baseline generation while maintaining comparable output quality. The system generates five-second video in 2.6 seconds, approaching real-time throughput.

This speedup exceeds prior acceleration methods by approximately a factor of two. The authors compare against CosVid and Self-Forcing, both of which achieve approximately 140x acceleration but exhibit quality degradation over longer sequences. DiagDistill maintains consistency over sequences of up to five minutes, suggesting robustness under the sustained operation our threat model requires.

The relevance to our feasibility classes is direct. Our original "edge full video synthesis" class assumed 12-month feasibility based on projected optimization trajectories extrapolated from TurboDiffusion (100 to 200x speedup) and SnapGen-V (mobile five-second generation in five seconds) [1, Section

2.2]. DiagDistill surpasses the TurboDiffusion benchmark within eight weeks of our publication, suggesting the aggressive scenario in our sensitivity analysis [1, Section 3.5] may better reflect the actual trajectory than the moderate scenario we treated as our baseline.

The technique has been tested on NVIDIA hardware with at least 24GB VRAM and 64GB RAM. This exceeds current wearable-class hardware but is within reach of phone-tethered configurations (AR glasses streaming to a high-end phone) and is consistent with the next generation of mobile SoCs expected within 12 to 18 months.

2.3 Mobile 3D Rendering: MobileGS

MobileGS [4] enables rendering of 3D Gaussian splatting scenes on mobile phones at over 120 frames per second using a Snapdragon 8 Gen 3 processor. The model has been compressed to 4.8 megabytes while preserving visual quality comparable to desktop implementations.

While MobileGS renders pre-reconstructed 3D scenes rather than generating new ones, its relevance to our threat model is twofold.

First, it demonstrates that real-time neural rendering on phone-class hardware is no longer aspirational; it is deployed and benchmarked. Our original feasibility analysis treated mobile neural rendering as a dependency for edge-only configurations. That dependency is now partially satisfied.

Second, the compression techniques used (shrinking from desktop-scale models to 4.8MB while preserving quality) represent advances in model distillation that transfer across applications. The same optimization principles applied to 3D rendering can be (and are being) applied to generative models, diffusion inference, and person segmentation.

2.4 Contextual Development: InSpatio-WorldFM

We note but do not incorporate into our revised feasibility estimates the release of InSpatio-WorldFM [5], a real-time generative frame model for spatial intelligence that generates multi-view consistent 3D environments from single images. WorldFM achieves interactive frame rates on a single RTX 4090 consumer GPU and explicitly targets edge deployment through model distillation.

WorldFM is relevant to our broader analysis of perceptual modification threats (addressed in companion papers) but does not directly impact the nudification pipeline, which operates on person-specific synthesis rather than environmental generation. We mention it here to document that spatial generation capabilities are advancing on the same trajectory as the capabilities more directly relevant to our threat model.

3. Revised Feasibility Assessment

3.1 Updated Pipeline

Incorporating the developments above, the revised pipeline latency estimates are:

Stage	Original Estimate	Revised Estimate	Change Factor
Camera input	<1ms	<1ms	Unchanged

Person detection + segmentation	35-60ms	15-30ms	MatAnyone2 quality at equivalent speed
Region crop + pose estimation	20-30ms	15-25ms	Improved segmentation reduces post-processing
Cloud upload (if applicable)	50-100ms+	50-100ms+	Network-dependent, unchanged
Synthesis model inference	100-500ms	40-200ms	DiagDistill trajectory applied
Video interpolation/consistency	100ms-5s	50ms-2s	DiagDistill long-sequence stability
AR display composition	5-10ms	5-10ms	Unchanged

The compression is most significant in Stages 5 and 6 (synthesis and consistency), where DiagDistill-class acceleration reduces the bottleneck that dominated our original latency analysis.

3.2 Revised Feasibility Classes

Class	Original Timeline	Revised Timeline	Basis
Cloud-assisted real-time	NOW	NOW	Unchanged
Edge static + interpolation	NOW	NOW	MatAnyone2 improves quality
Edge full video synthesis	12 months	6-9 months	DiagDistill exceeds projected speedup
Edge real-time (<500ms)	24-36 months	18-24 months	MobileGS + DiagDistill trajectory

3.3 Sensitivity Analysis Revision

Our original analysis provided conservative, moderate, and aggressive scenarios [1, Section 3.5]. The observed development trajectory most closely matches our original aggressive scenario, which assumed breakthrough optimization. We now treat our original aggressive scenario as the revised moderate scenario and propose a new aggressive scenario:

Revised Aggressive Scenario (assumes continued trajectory): - Edge full video synthesis: 3 to 6 months - Edge real-time: 12 to 18 months

We emphasize that this is a trajectory-based projection, not a guarantee. Hardware constraints, thermal limitations on wearable form factors, and the difference between benchmark performance and sustained operation introduce uncertainty. However, the direction of compression is unambiguous.

4. Implications

4.1 Structural Mismatch Widened

Our original analysis identified a structural mismatch: the threat matures faster than legislative and regulatory processes can respond [1, Section 7.3]. The timeline compression documented here widens that mismatch.

If edge full video synthesis is achievable in 6 to 9 months rather than 12, and if legislative processes require 3 to 5 years (a timeline documented in prior technology policy analyses), the gap between capability and governance has expanded. The window for proactive policy response has narrowed correspondingly.

4.2 Open-Source Deployment as Acceleration Factor

All three developments documented in Section 2 share a critical property: they are fully open source with deployment-ready code, documentation, and in some cases hosted demo environments. None require technical sophistication to access or integrate.

This pattern confirms our original observation that the barrier to ambient synthesis is not capability development but integration [1, Section 7.1]. The components exist. They are free. They are documented. The remaining barrier is glue code connecting them into a pipeline, and that barrier is precisely the task at which current large language models excel.

4.3 Quality Convergence

A subtler implication of MatAnyone2's segmentation quality is that the visual quality gap between cloud-assisted and edge-only configurations is narrowing. Our original analysis noted that cloud-assisted configurations produce higher quality output [1, Section 3.3]. Improved segmentation at the edge means the edge pipeline produces cleaner input to the synthesis stage, reducing the quality penalty of local processing. This matters because edge-only operation is more threatening: it leaves no network trace, requires no external service, and is fully deniable.

5. Updated Recommendations

Our original recommendations [1, Section 8] remain applicable. We add the following:

First, the compressed timeline strengthens the case for immediate legislative action targeting creation rather than distribution. Every month of delay brings the edge-only deployment class closer to feasibility, at which point enforcement becomes substantially more difficult.

Second, model-hosting platforms (HuggingFace, CivitAI, GitHub) should develop policies addressing not only individual model harms but pipeline composition. Hosting a person segmentation model, a synthesis acceleration technique, and a mobile rendering optimizer separately does not constitute harm. But the combination of all three within an integrated pipeline may. Platform policies that evaluate models only in isolation cannot detect composition risks. We note this as an instance of a broader governance gap addressed in our companion methodology paper [6].

Third, the acceleration rate itself has policy implications. Eight weeks between our original publication and the developments that compress our timeline suggests that feasibility projections in this domain have a short shelf life. Governance frameworks that depend on static risk assessments will perpetually lag. Continuous monitoring of capability developments and their combinatorial implications is necessary.

6. Conclusion

The developments reported here do not change the nature of the threat we identified in [1]. They change its timing. Cloud-assisted ambient non-consensual synthesis remains feasible today. Edge-only synthesis has moved closer by several months across our feasibility classes. The gap between technical capability and governance response has widened.

The most significant finding is not any individual tool but the rate at which independently developed capabilities converge on the dependencies we identified. Eight weeks produced a 270x inference speedup, a 140MB state-of-the-art segmentation model, and a 4.8MB mobile 3D renderer. None of these teams cited our threat analysis. None appear aware of the pipeline they are collectively enabling. This is exactly the pattern our original analysis predicted and exactly the pattern that current governance frameworks are not designed to detect.

Acknowledgements

This paper was developed using speech-to-text transcription and Claude (Anthropic) as assistive technology accommodations for ADHD and ASD. These tools were used for transcription of verbal ideation, organizational formatting, and structural editing. All research questions, arguments, analytical contributions, methodology, and conclusions are solely the author's own. No AI system generated the intellectual content, theoretical frameworks, or original analysis presented in this work.

References

[1] T. Gilly, "Ambient Non-Consensual Image Synthesis: A Convergence Threat Analysis of AR Wearables, Real-Time Video Generation, and Open-Source Nudification Models," Zenodo, Jan. 2026. https://doi.org/10.5281/zenodo.18297286

[2] P. Yang, S. Zhou, K. Hao, and Q. Tao, "MatAnyone 2: Scaling Video Matting via a Learned Quality Evaluator," in Proc. CVPR, 2026. arXiv:2512.11782. GitHub: https://github.com/pq-yang/MatAnyone2. HuggingFace Demo: https://huggingface.co/spaces/PeiqingYang/MatAnyone2.

[3] J. Liu, X. Liu, K. Mei, Y. Wen, M.-H. Yang, and W. Liu, "Streaming Autoregressive Video Generation via Diagonal Distillation," in Proc. ICLR, 2026. arXiv:2603.09488. GitHub: https://github.com/Sphere-AI-Lab/diagdistill. Project page: https://spherelab.ai/diagdistill/.

[4] X. Du et al., "Mobile-GS: Real-time Gaussian Splatting for Mobile Devices," in Proc. ICLR, 2026. arXiv:2603.11531. Project page: https://xiaobiaodu.github.io/mobile-gs-project/.

[5] InSpatio-WorldFM: An Open-Source Real-Time Generative Frame Model for Spatial Intelligence. arXiv:2603.11911, March 2026.

[6] T. Gilly, "Emergent Convergence Risk: Why Existing AI Governance Cannot Detect Combinatorial Threats and a Proposed Methodology," forthcoming, 2026.

Word Count: Approximately 2,200 words (excluding references)