EMERGENT CONVERGENCE RISK · COMBINATORIAL THREATS · THE STACK · GOVERNANCE BLIND SPOT · CONVERGENCE SCANNING · SYSTEMIC RISK

The Accidental Stack

How eight tools released in a single week became the parts list for a population-scale perceptual control system that no developer intended, designed, or recognized.

Travis Gilly · Real Safety AI Foundation

The Accidental Stack: A Case Study in Emergent Perceptual Control Infrastructure from Independently Developed AI Capabilities · Working paper, 2026

8

OPEN-SOURCE TOOLS
RELEASED IN SEVEN DAYS

One weekly roundup video covered roughly fifteen new releases. A subset of them was the component stack for continuous, real-time modification of a person's experience of physical reality.

Each tool was presented individually, with a demonstration and a link to open code. No evaluation asked what any tool enables when combined with the others.

SOURCES Gilly, The Accidental Stack (2026), secs. 1-2; AI Search, AI NEWS roundup (March 2026).

NO ONE DESIGNED IT

The system is emergent. It exists in the spaces between tools, at the integration surfaces, and nowhere else.

Existing AI governance evaluates capabilities in isolation; it is therefore structurally incapable of detecting emergent convergence threats. The gap is not an oversight in implementation. It is a limit of the methodological paradigm.

SOURCES Gilly, The Accidental Stack (2026), secs. 1, 7.

Each chemical sits below its flash point. The inspector checks them one by one and stamps the manifest compliant.

Stored adjacent, they form a reactive mixture. The warehouse passes every inspection and burns anyway. The inspector was not negligent; the inspection method was inadequate to the risk class.

> *AI governance is currently running chemical-by-chemical inspections on a warehouse full of reactive capabilities.*

SOURCES Gilly, The Accidental Stack (2026), sec. 1.

PART ONE

The Stack

Eight capabilities, each individually benign.

I

Reviewed in isolation, every tool reads as a legitimate advance in vision, generation, rendering, or speech. The danger appears only at the seams.

MatAnyone2 SEGMENT

Real-time video matting that separates a person from any background. Provides real-time isolation of individuals for downstream processing.

MobileGS RENDER

A Gaussian-splatting renderer at 4.8 MB hitting 120-plus FPS on a phone. Provides 3D scene rendering on mobile and wearable-class hardware.

InSpatio-WorldFM GENERATE

A real-time generative frame model producing spatially consistent 3D environments with persistent memory. Provides coherent alternative worlds to overlay reality.

Diagonal Distillation CATALYZE

A 270x video-generation speedup, five seconds of video in 2.6 seconds. Generates nothing itself; converts laboratory demos into continuous, deployed systems.

SOURCES Yang et al. (2026a); Zhang, X., et al. (2026); Du et al. (2026); Liu et al. (2026), all in Gilly, The Accidental Stack (2026), sec. 2.1.

TaDa CLONE

Voice cloning from ten seconds of reference audio, with the fastest synthesis among benchmarked competitors.

Effect Maker TRANSFER

Copies specific visual effects from one video and applies them to new subjects.

Fish Audio S2 MODULATE

Speech synthesis with inline tags for pausing, emphasis, laughter, whispering, and shouting.

Shotverse COMPOSE

Multi-shot cinematic video with smooth camera transitions and consistent characters across cuts.

> **Already shipping, not in the video:** Google Maps Gemini immersive navigation, delivered to billions of phones, and RL3D Edit, text-prompt editing of objects and characters inside complete 3D scenes.

SOURCES Gilly, The Accidental Stack (2026), secs. 2.1-2.2; Daniel (2026); Wang et al. (2026).

A surface is any interface where two tools connect. At each one, the question is the same: who gets hurt?

SURFACE 1

Extract + Replace

Segment a person, regenerate everything around them. The person becomes the only real element in a synthetic scene.

SURFACE 3

Clone + Modulate

Generate emotionally modulated speech in anyone's cloned voice, controllable word by word, in real time.

THE CATALYST

Acceleration

The accelerator harms no one alone. It removes the latency barrier that bounded every other tool's harm, and makes the stack operational.

SOURCES Gilly, The Accidental Stack (2026), sec. 3 (integration surface analysis).

A system that continuously renders an alternative reality.

visually coherent | spatially persistent | auditorily complete | emotionally adaptive | cinematically smooth

Every person in the user's environment can be visually modified or replaced. Every surface, object, and space can be regenerated. Every voice can be substituted. No individual developer built this; no individual tool evaluation would detect it.

SOURCES Gilly, The Accidental Stack (2026), sec. 3.2.

PART TWO

Why Governance Misses This

Three instruments, one blind spot.

II

Each evaluates AI capabilities as discrete, bounded entities. That is appropriate for the risks they were built for, and inadequate for a threat that lives in the relationships between entities.

EU AI Act

RISK TIERS BY PURPOSE

Every tool here would likely be classed minimal or limited risk on its own. There is no mechanism for evaluating what minimal-risk tools become once composed.

NIST AI RMF

GOVERN, MAP, MEASURE, MANAGE

It needs a system to evaluate. Here there is none yet: nothing deployed to map, nothing built to measure. It was designed for systems someone intends to deploy.

MIT Risk Repository

A CATALOG OF KNOWN RISKS

A risk must be recognized by a researcher before it can be listed. It maps explored territory; it is not a method for exploring new territory.

SOURCES European Parliament & Council (2024); NIST (2023); MIT FutureTech (2024); Gilly, The Accidental Stack (2026), sec. 4.

The periodic table catalogs elements. Safety sheets describe single chemicals. Neither tells you what happens when you combine them.

That requires a different discipline: reaction chemistry, the study of interactions between substances rather than the substances themselves.

> *AI governance needs its reaction chemistry.*

SOURCES Gilly, The Accidental Stack (2026), sec. 4.4.

PART THREE

Toward Convergence Scanning

No systematic method exists yet. Here is what one would require.

III

There is no systematic way to scan independently developed capabilities for emergent harm at integration surfaces. The paper does not claim to solve this; it specifies the shape of a solution.

1 **Operate on capabilities, not products.** The threat exists before anyone assembles a system.

2 **Be continuous.** New capabilities ship daily; a point-in-time audit is stale immediately.

3 **Scale across domains.** A single-domain scanner misses cross-domain convergence by definition.

4 **Tolerate ambiguity.** Filter a vast combination space without too many false positives or negatives.

5 **Be grounded in harm.** Ask whether the combination opens a new path to harm, not whether it is more capable.

SOURCES Gilly, The Accidental Stack (2026), sec. 5.2.

HARM BLINDNESS, AT THE SEAMS

Move the question from checkpoints inside one product to surfaces between independent ones.

For each compatible pair, ask: what does the combination enable that neither does alone? Who is newly affected? How reversible is the effect? What consent exists for them? And what is the minimum barrier to achieving it?

THE DECIDING QUESTION

If the join needs a research team, the risk is bounded. If it needs glue code a model writes in minutes, the risk is immediate.

Every tool in this case study is open, documented, and format-compatible. The glue code is trivial.

SOURCES Gilly, The Harm Blindness Framework (2025-2026); Gilly, The Accidental Stack (2026), sec. 5.3.

Cyberpunk 2077

Control through invasive hardware. Centralized because the technology is centralized. There is an Arasaka to regulate.

We Happy Few

Control through a self-administered drug. Decentralized, but there is still a factory you could shut down.

The reality

Emergent, distributed, invisible. No corporation, no factory. Thousands of repositories, each passing every individual safety check.

Fiction assumed perceptual control would be intentional, centralized, and recognizable. The reality is accidental, distributed, and invisible to every existing method.

SOURCES CD Projekt Red (2020); Compulsion Games (2018); Gilly, The Accidental Stack (2026), sec. 6.

IN CONCLUSION

The question is not whether this stack is dangerous. It is whether governance will learn to scan for convergence, or keep inspecting chemicals one at a time while the warehouse fills.

The components are available, open source, documented, and trivially combinable, and no systematic process exists to detect that fact. If this is the only dangerous convergence in the current landscape, we got extraordinarily lucky. The paper does not believe we did.

SOURCES Gilly, The Accidental Stack (2026), sec. 7.

Gilly, T. (2026). *The Accidental Stack: A Case Study in Emergent Perceptual Control Infrastructure from Independently Developed AI Capabilities.* Real Safety AI Foundation, working paper.

AI Search. (2026, March). AI maps, realtime 3D worlds, multi-shot videos, new TTS, new anime model: AI NEWS [Video]. YouTube.

CD Projekt Red. (2020). Cyberpunk 2077 [Video game]. CD Projekt.

Compulsion Games. (2018). We Happy Few [Video game]. Gearbox Publishing.

Daniel, M. (2026, March 12). How we're reimagining Maps with Gemini. Google Blog. https://blog.google/products-and-platforms/products/maps/ask-maps-immersive-navigation/

Du, X., Jiang, Y., Ye, T., Han, S., & Liu, S. (2026). Mobile-GS: Real-time Gaussian splatting for mobile devices. In Proceedings of the International Conference on Learning Representations (ICLR). arXiv:2603.11531. https://xiaobiaodu.github.io/mobile-gs-project/

European Parliament & Council. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act). Official Journal of the European Union. EUR-Lex: 32024R1689.

Gilly, T. (2025-2026). The Harm Blindness Framework: A practical application methodology for stakeholder harm prevention in technology development. Real Safety AI Foundation. https://realsafetyai.org/framework

Gilly, T. (2026b). We Happy Few: Therapeutic perceptual modification, regulatory pathway precedent, and the governance gap in cloud-rendered reality. Forthcoming.

Liu, J., Liu, X., Mei, K., Wen, Y., Yang, M.-H., & Liu, W. (2026). Streaming autoregressive video generation via diagonal distillation. In Proceedings of the International Conference on Learning Representations (ICLR). arXiv:2603.09488. https://github.com/Sphere-AI-Lab/diagdistill

MIT FutureTech. (2024). AI Risk Repository. Massachusetts Institute of Technology. https://airisk.mit.edu

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

Wang, J., et al. (2026). Geometry-guided reinforcement learning for multi-view consistent 3D scene editing. arXiv preprint arXiv:2603.03143. https://github.com/AMAP-ML/RL3DEdit

Yang, P., Zhou, S., Hao, K., & Tao, Q. (2026a). MatAnyone 2: Scaling video matting via a learned quality evaluator. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). arXiv:2512.11782. https://github.com/pq-yang/MatAnyone2

Zhang, X., et al. (2026). InSpatio-WorldFM: An open-source real-time generative frame model for spatial intelligence. arXiv preprint arXiv:2603.11911. https://inspatio.github.io/worldfm/