

Combating Falsification of Speech Videos with Live Optical Signatures

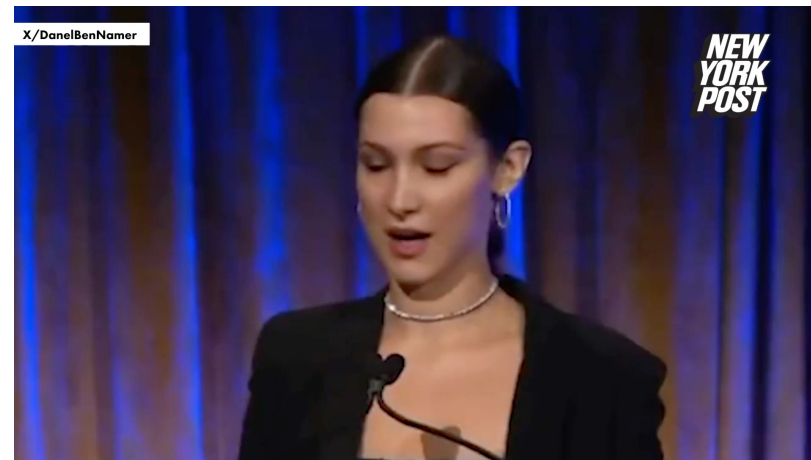
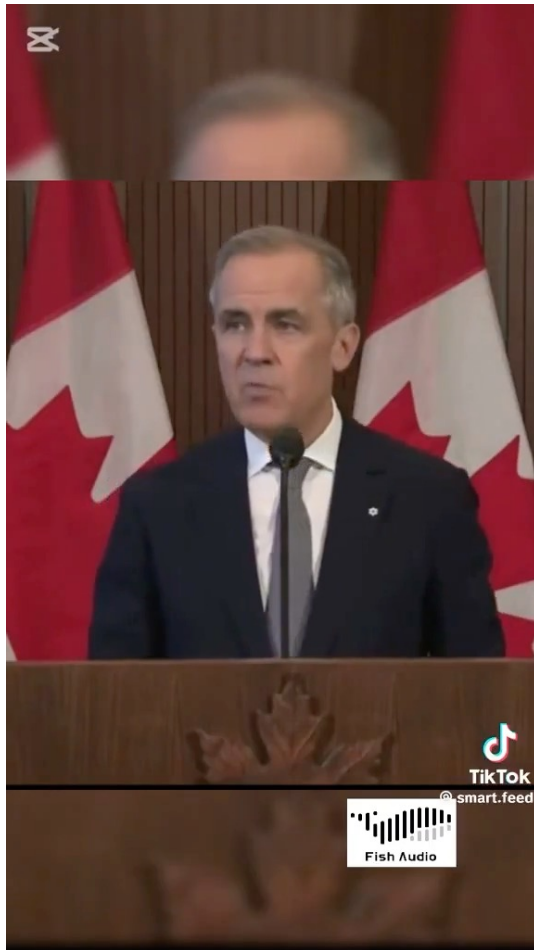
Hadleigh Schwartz¹, Xiaofeng Yan¹, Charles J. Carver², Xia Zhou¹

¹Columbia University

²Massachusetts Institute of Technology



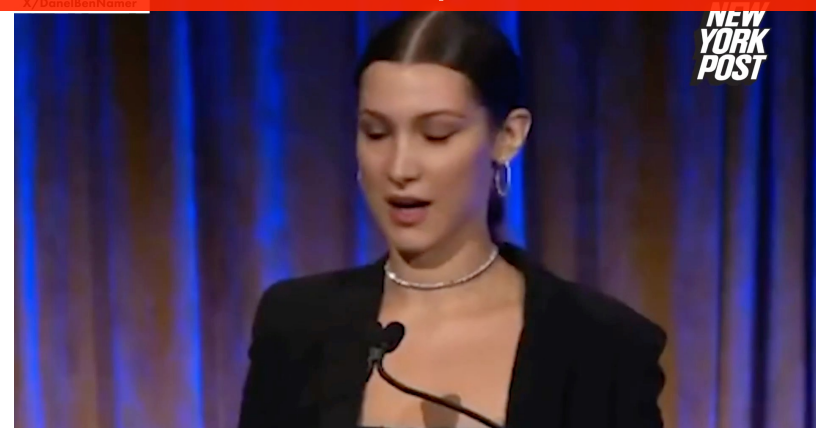
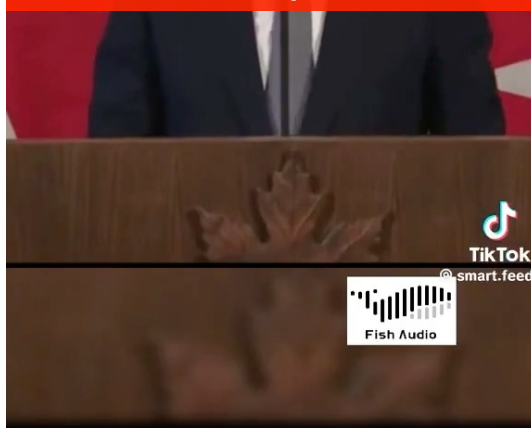
Unique Influence of Speech Videos



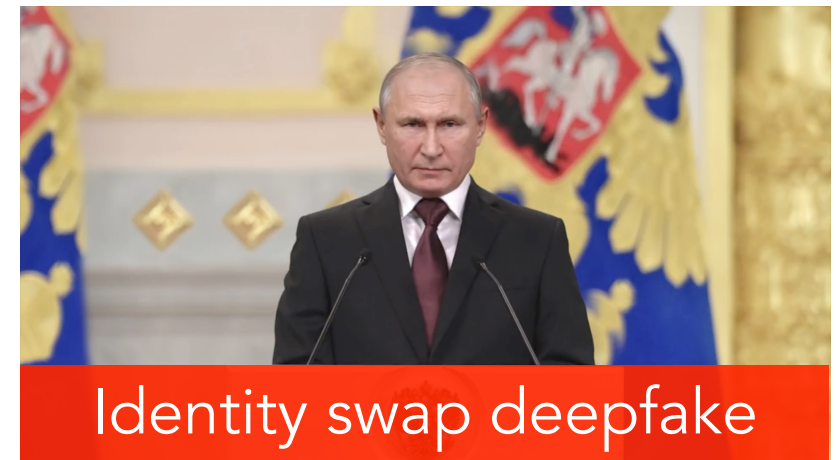
Unique Influence of Speech Videos



Lip/face reenactment deepfake

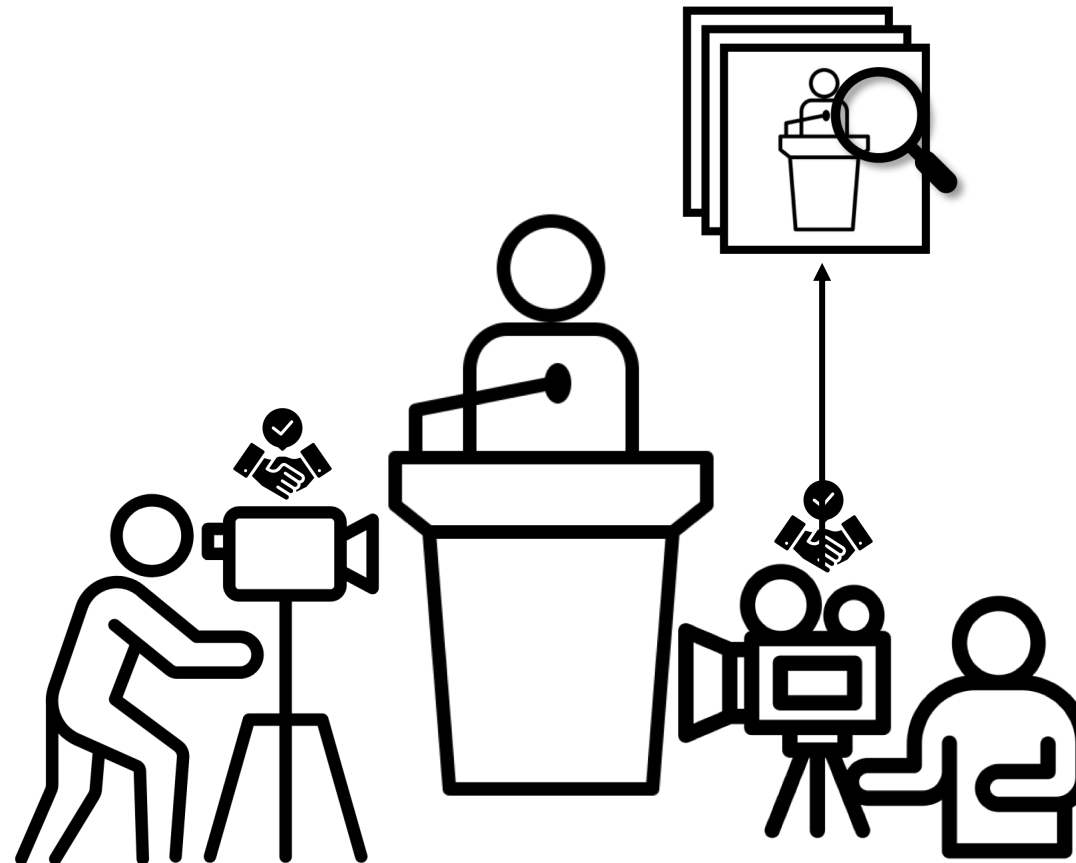


Slowed down



Identity swap deepfake

Existing Solutions



Passive detectors
Search for editing/
deepfake artifacts

✗ Difficult with AI
advances

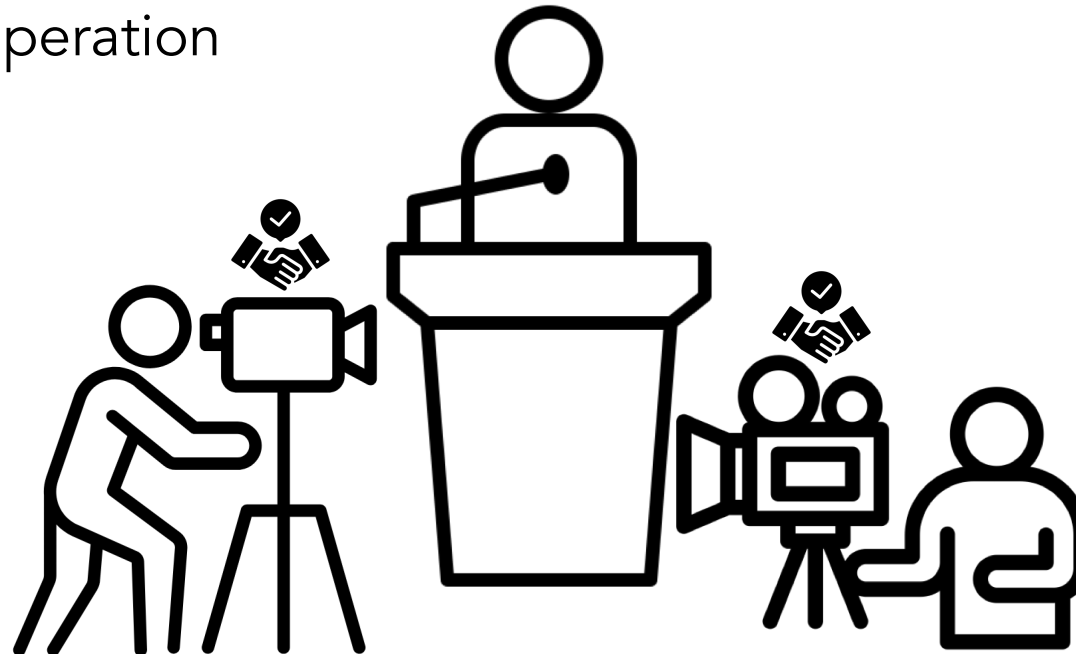
Existing Solutions

Proactive digital approaches
Digitally sign videos at creation,
watermark synthetic content

✗ Device & user cooperation

Passive detectors
Search for editing/
deepfake artifacts

✗ Difficult with AI
advances

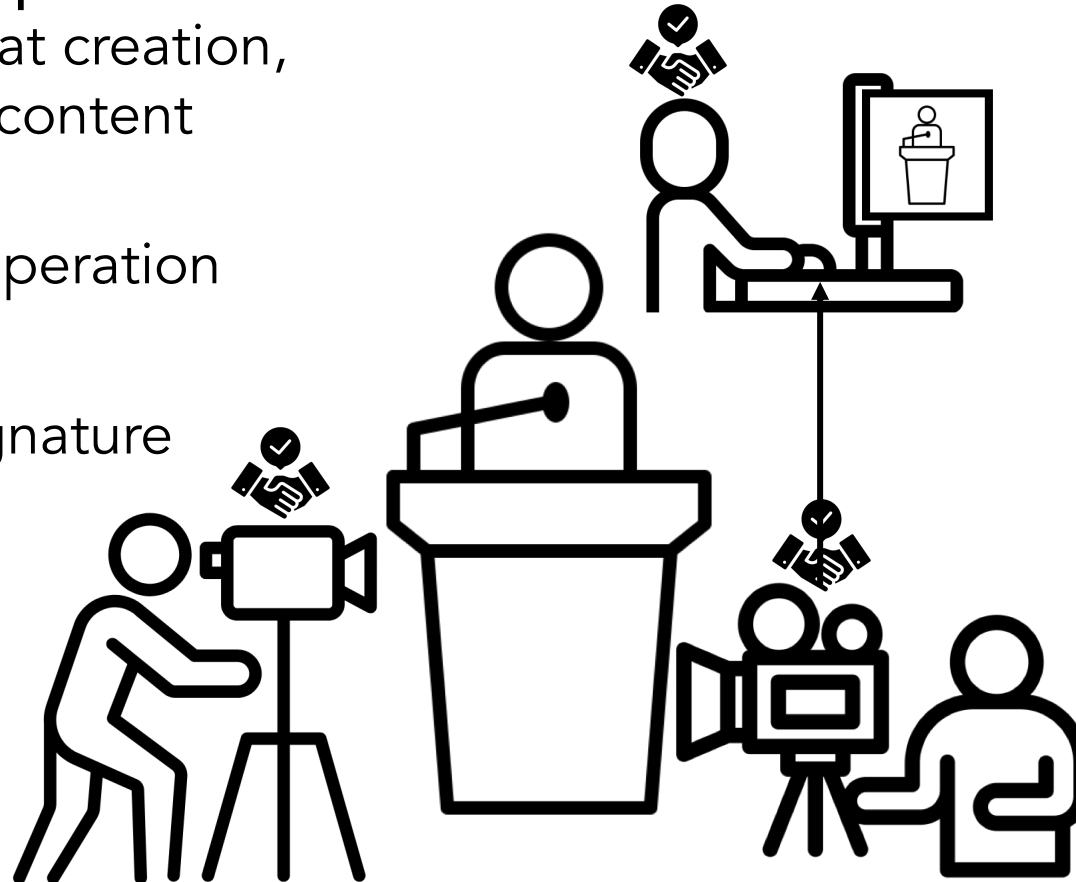


Existing Solutions

Proactive digital approaches
Digitally sign videos at creation,
watermark synthetic content

✗ Device & user cooperation

✗ Benign edits → signature
regeneration

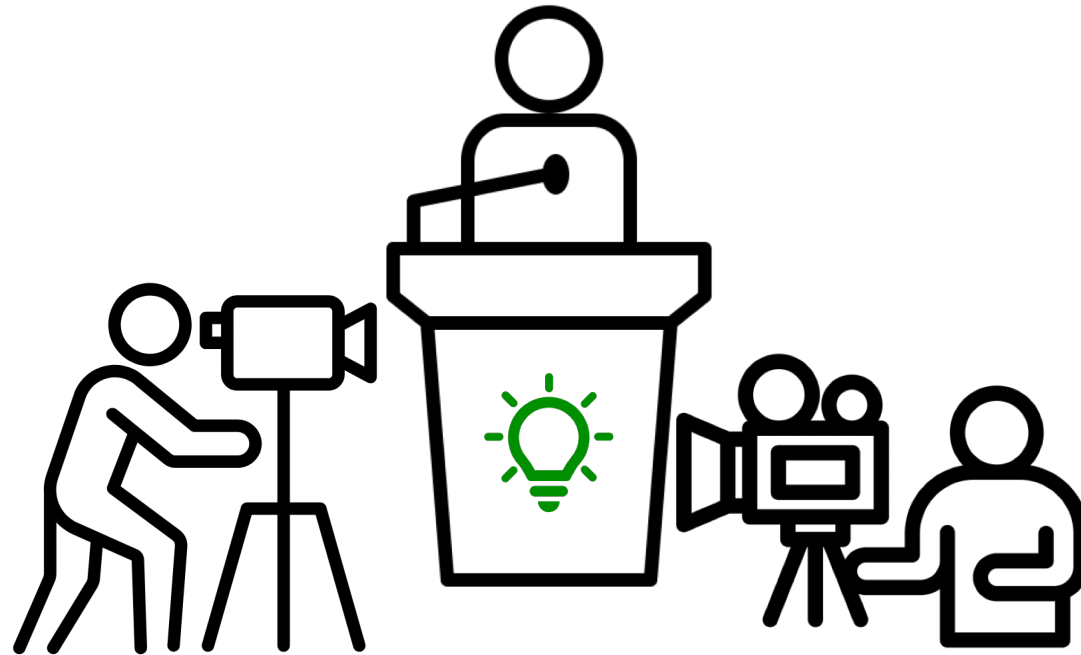


Passive detectors
Search for editing/
deepfake artifacts

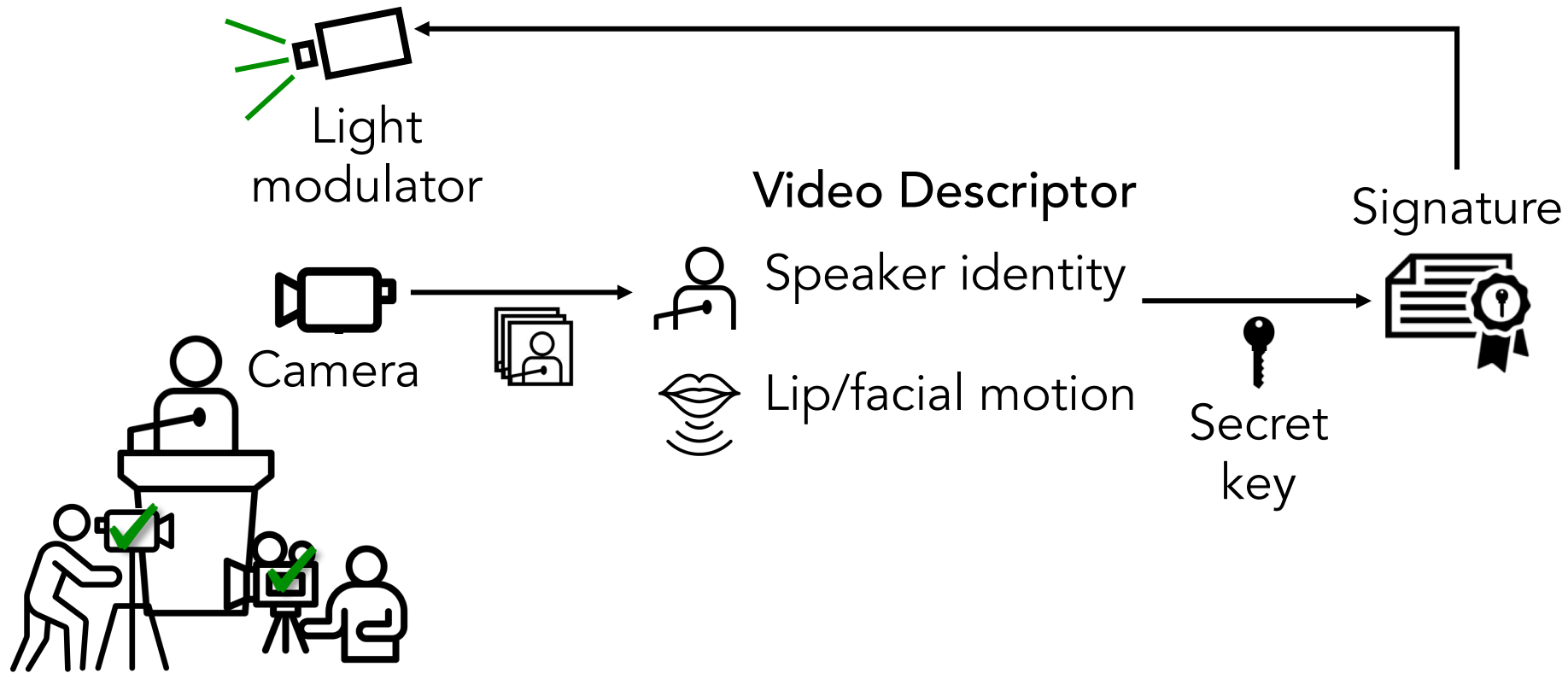
✗ Difficult with AI
advances

Our Approach

Create signatures physically at the live speech site using light

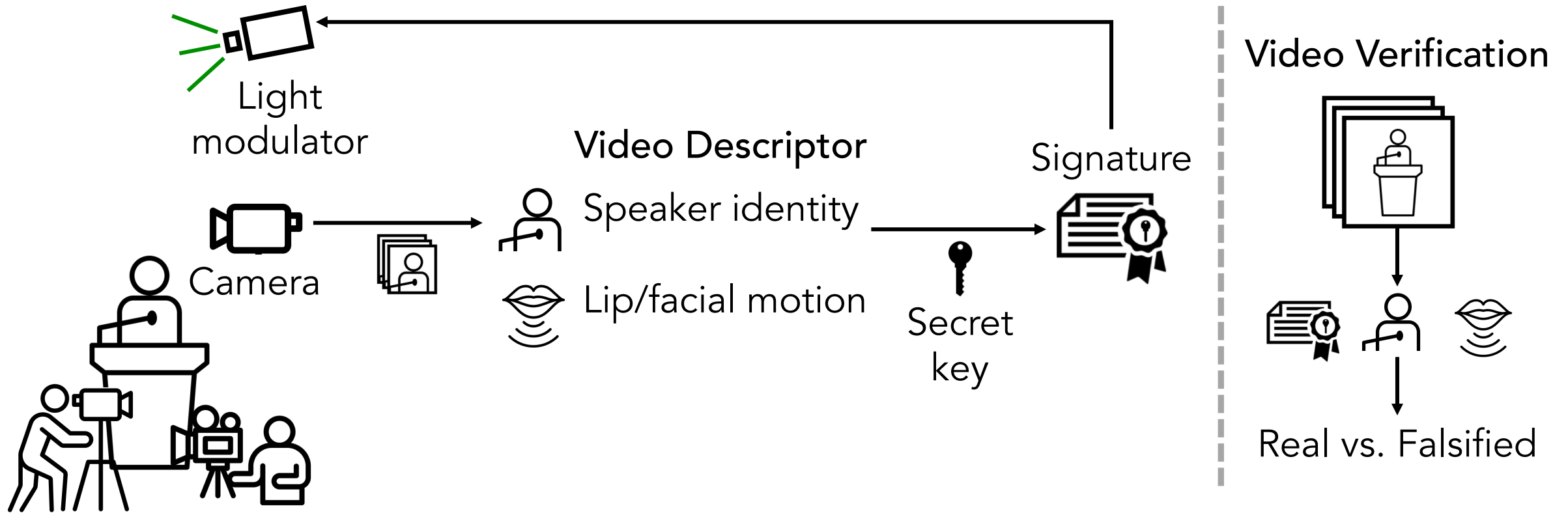


Physical Signatures with Light



✓ Naturally embedded into all real video recordings; no external cooperation required!

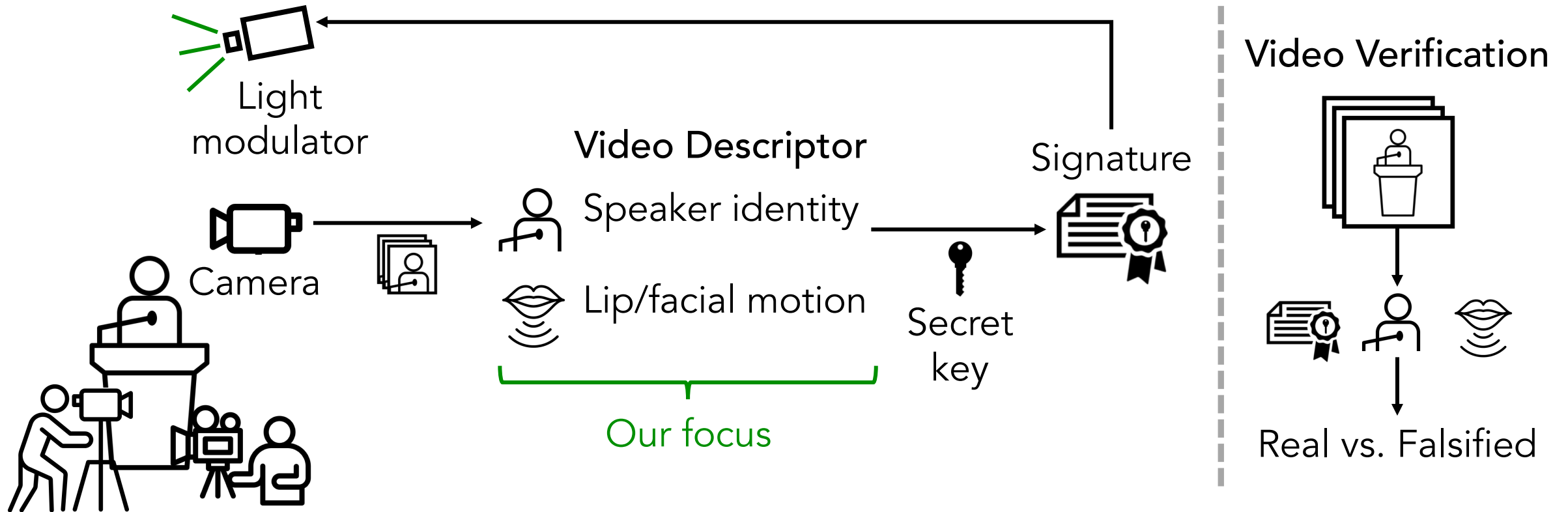
Physical Signatures with Light



✓ Naturally embedded into all real video recordings; no external cooperation required!

✓ Semantically-meaningful features robust to benign edits

Physical Signatures with Light



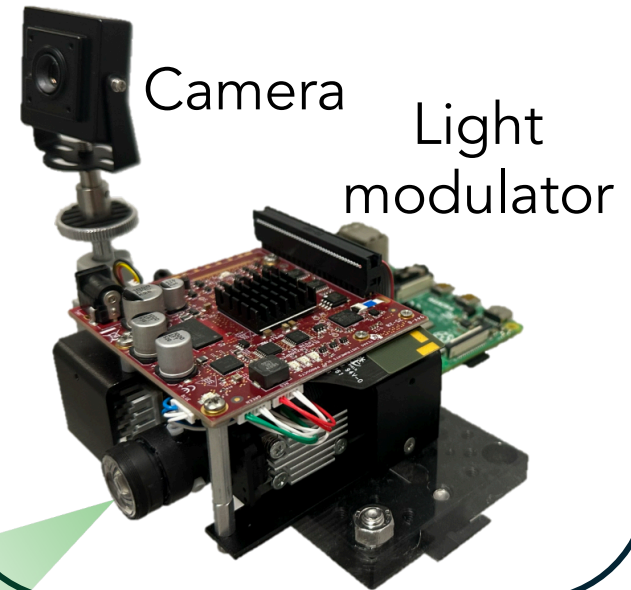
✓ Generic framework for protecting speech videos

VeriLight

Physical signatures using *imperceptible* modulated light

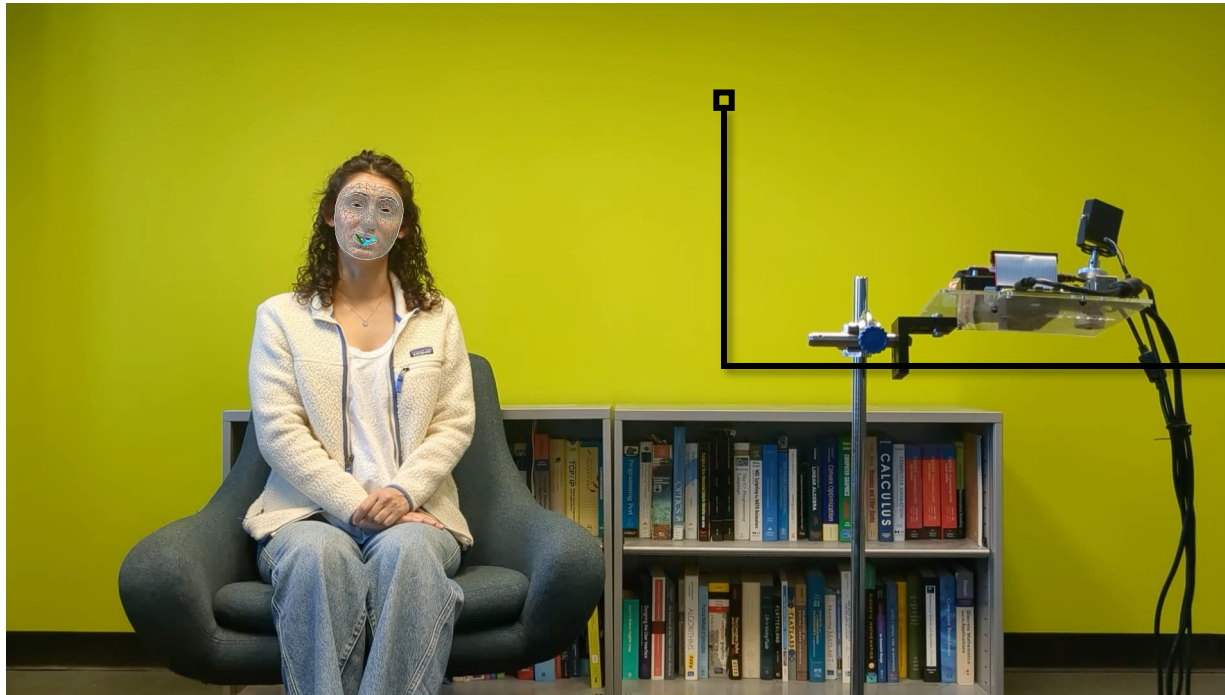


VeriLight Core Unit



VeriLight

Physical signatures using *imperceptible* modulated light



Embedded signature signals



VeriLight

Physical signatures using *imperceptible* modulated light



✓ Verified

VeriLight

Physical signatures using *imperceptible* modulated light



✗ Lip/facial motion
falsification detected

VeriLight

Physical signatures using *imperceptible* modulated light



✗ Identity falsification
detected

Threat Model

Video Producers

Create legitimate videos



Record with any standard RGB camera



60°, 3m from speaker (further with zoom)



Benign edits (e.g., compression, filter)

Attackers

Create falsified videos



Access VeriLight algorithms



Extensive comp. resources



Cannot access secret key, tamper with hardware



ID + lip/face motion manipulations

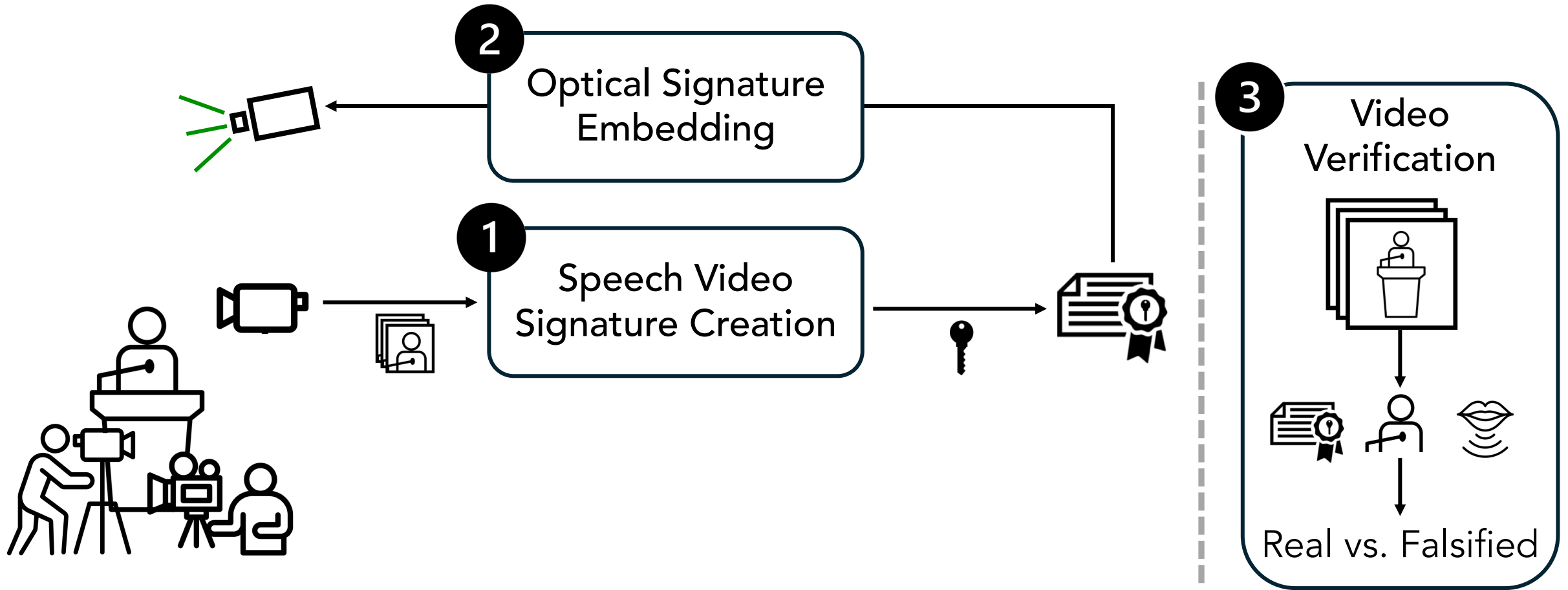


Malicious temporal edits (e.g., slowdown)

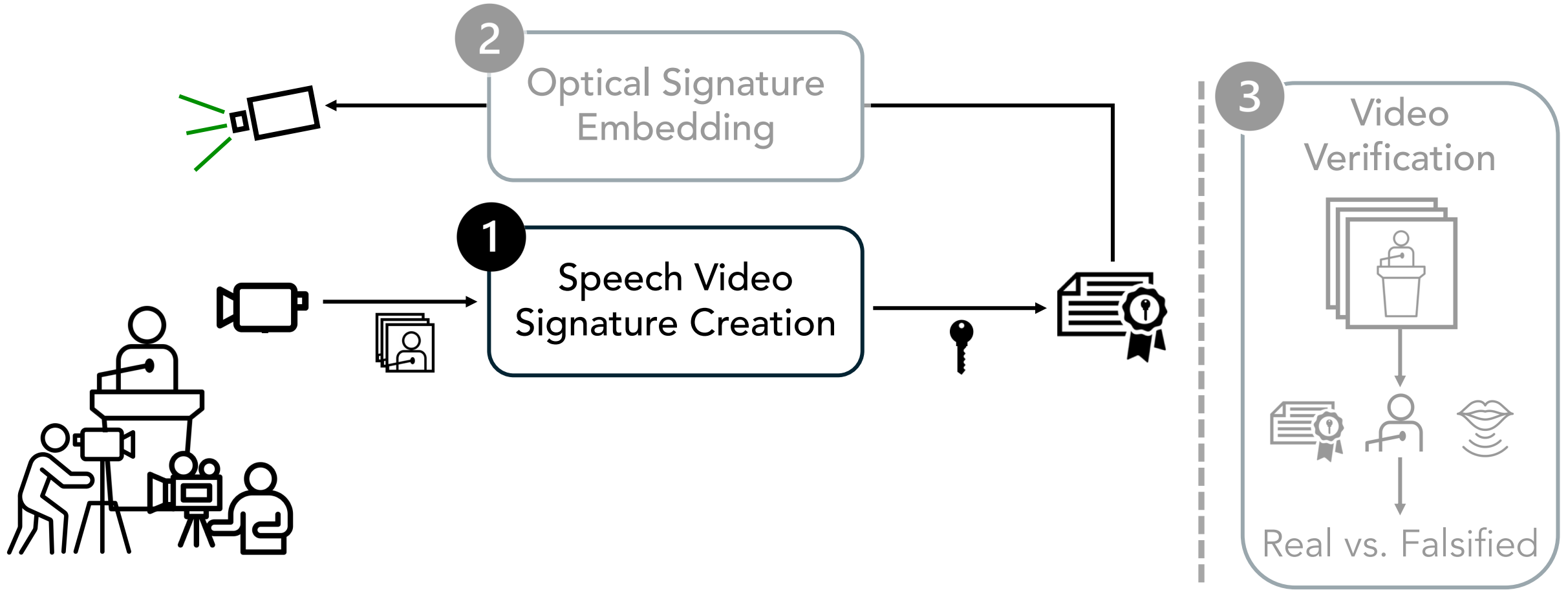


Copy or mimic signature signals

VeriLight Design Components



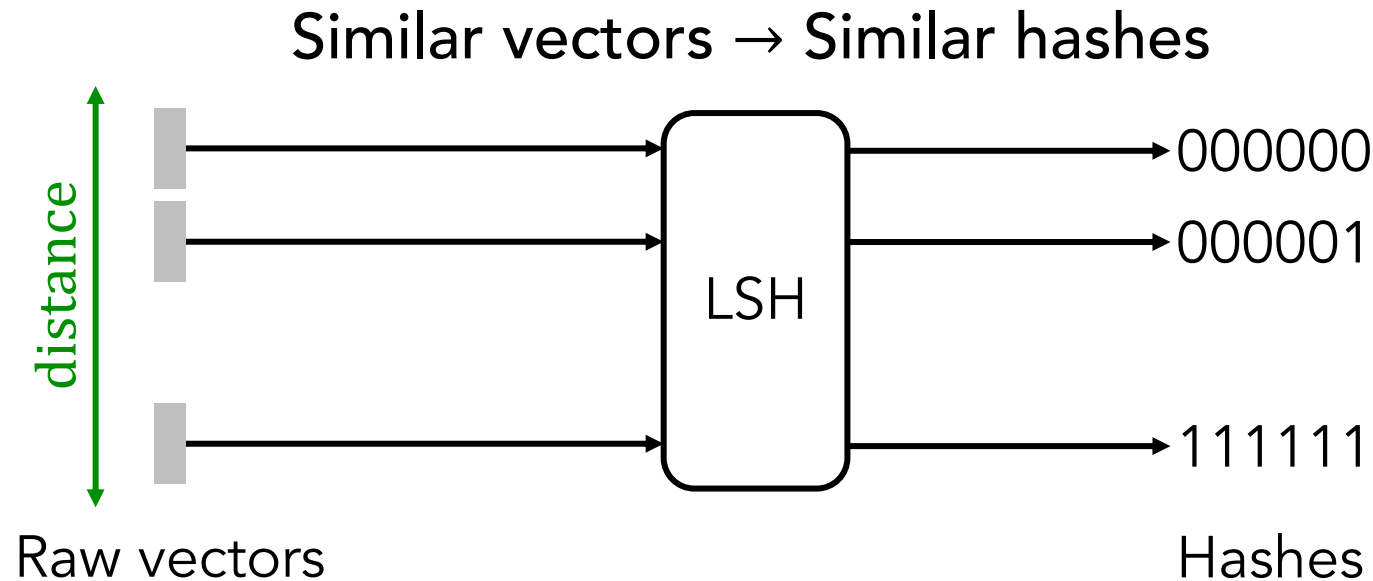
VeriLight Design Components



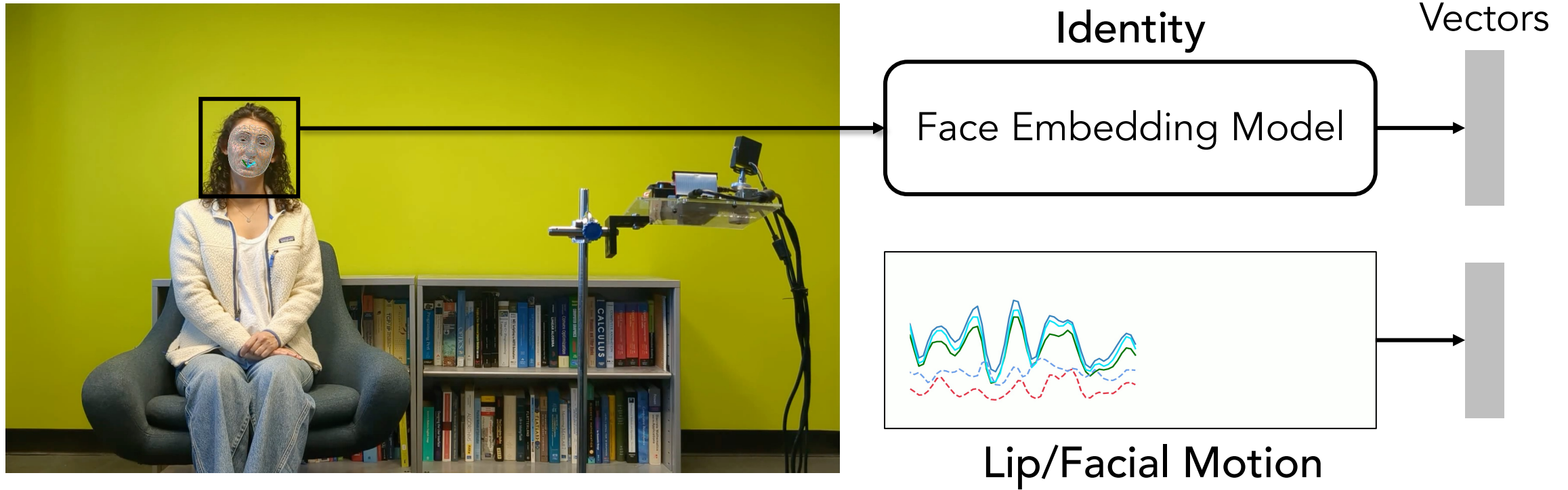
Speech Video Signature Creation

Challenge: **Highly-compact, pose-invariant** video descriptors

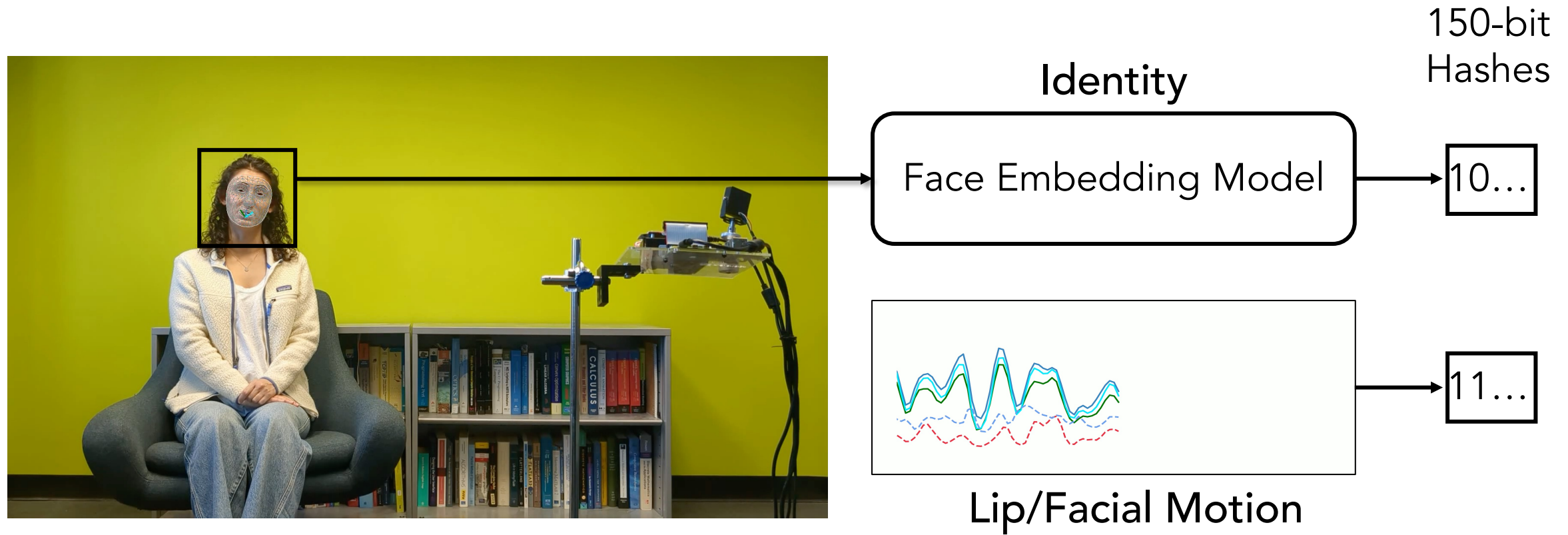
Solution: Video descriptor framework using locality-sensitive hashing (LSH)



Speech Video Signature Creation



Speech Video Signature Creation



Speech Video Signature Creation

Signature Assembly

Video Descriptor

Identity Feature Hash	Lip/Facial Motion Feature Hash
--------------------------	-----------------------------------

Speech Video Signature Creation

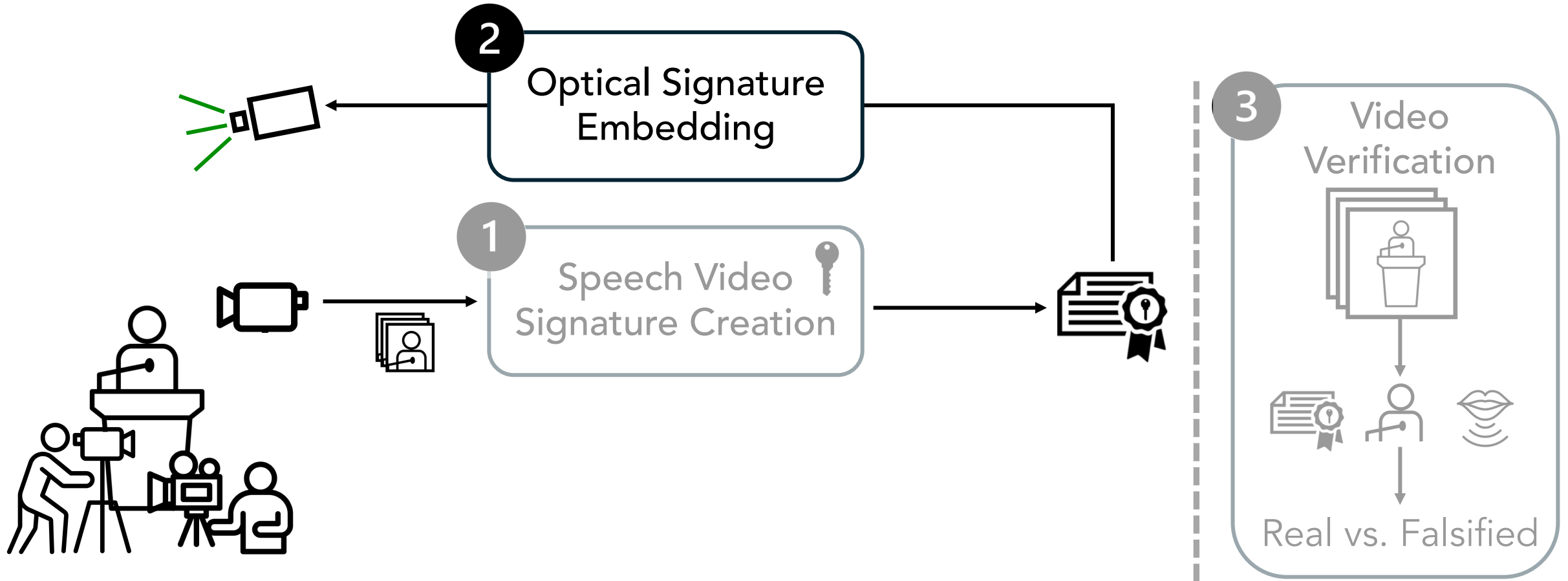
Signature Assembly

Video Descriptor

Identity Feature Hash	Lip/Facial Motion Feature Hash	Window #	Core Unit ID	Date	 HMAC
-----------------------	--------------------------------	----------	--------------	------	--

Signature

VeriLight Design Components

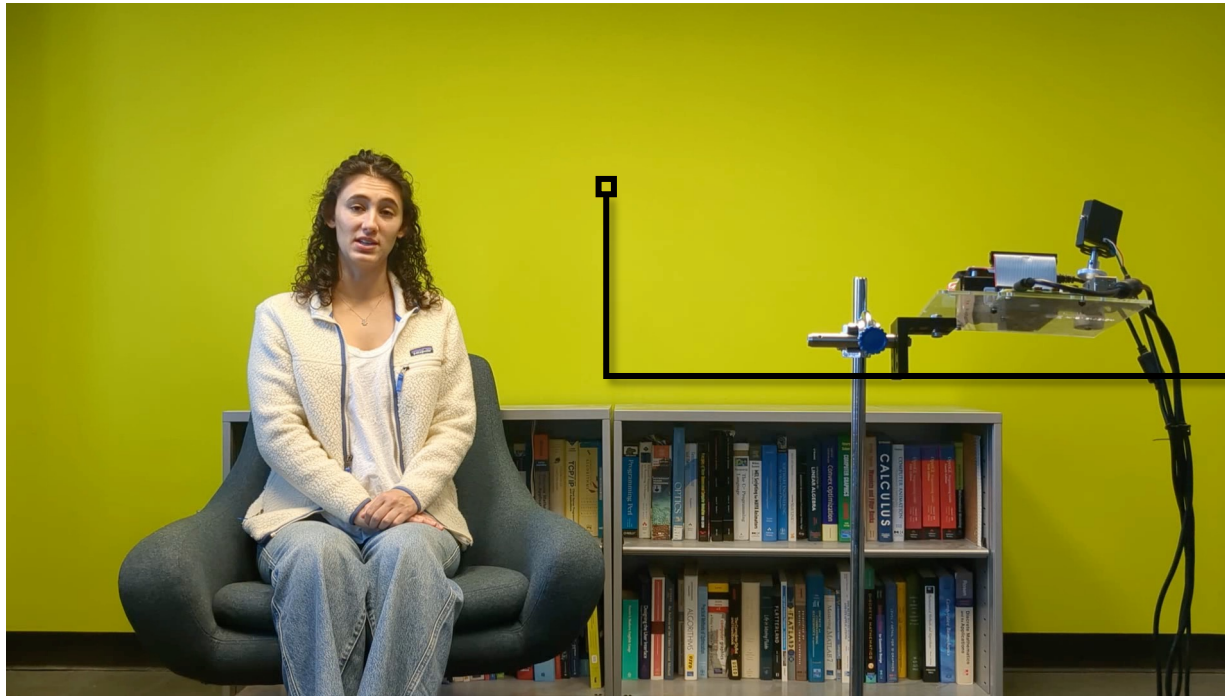


Optical Signature Embedding

Challenge: Balance bandwidth, post-processing/environment resilience, and imperceptibility

Optical Signature Embedding

Solution: Subtle spatial and temporal modulations of light

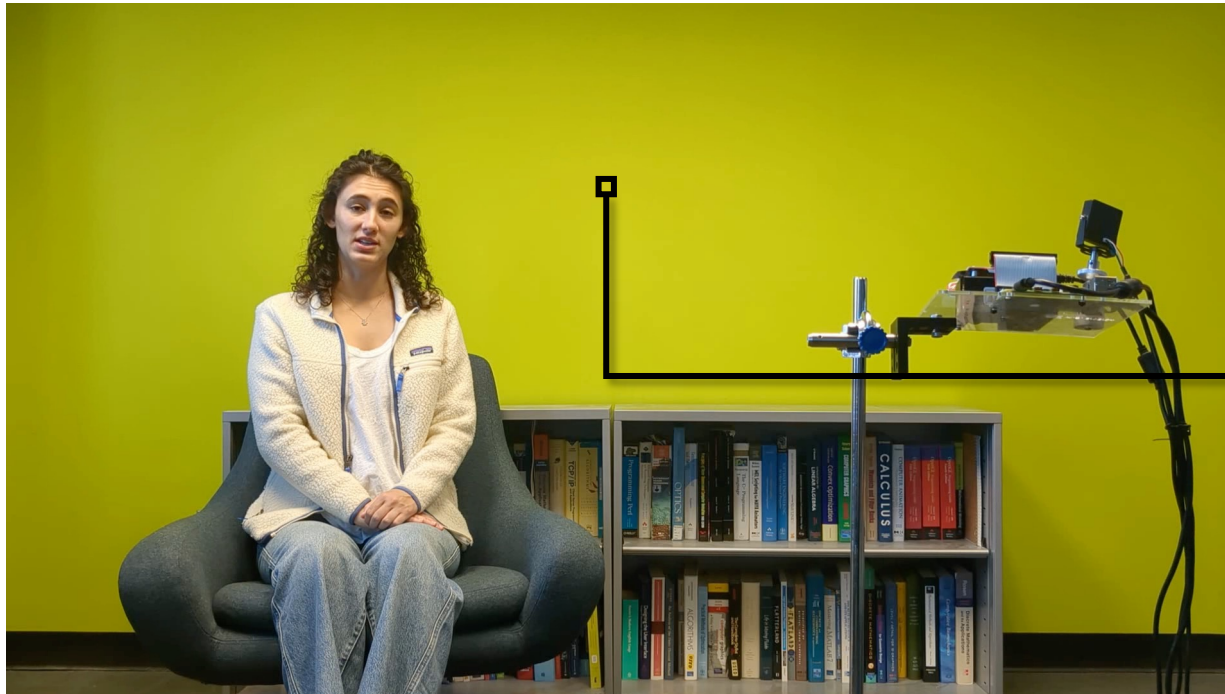


Video Pixel Intensity



Optical Signature Embedding

- ✓ >200 bps embedding bandwidth
- ✓ Imperceptible on *any* projection surface



Video Pixel Intensity



Optical Signature Embedding

✓ Environment and post-processing resilience

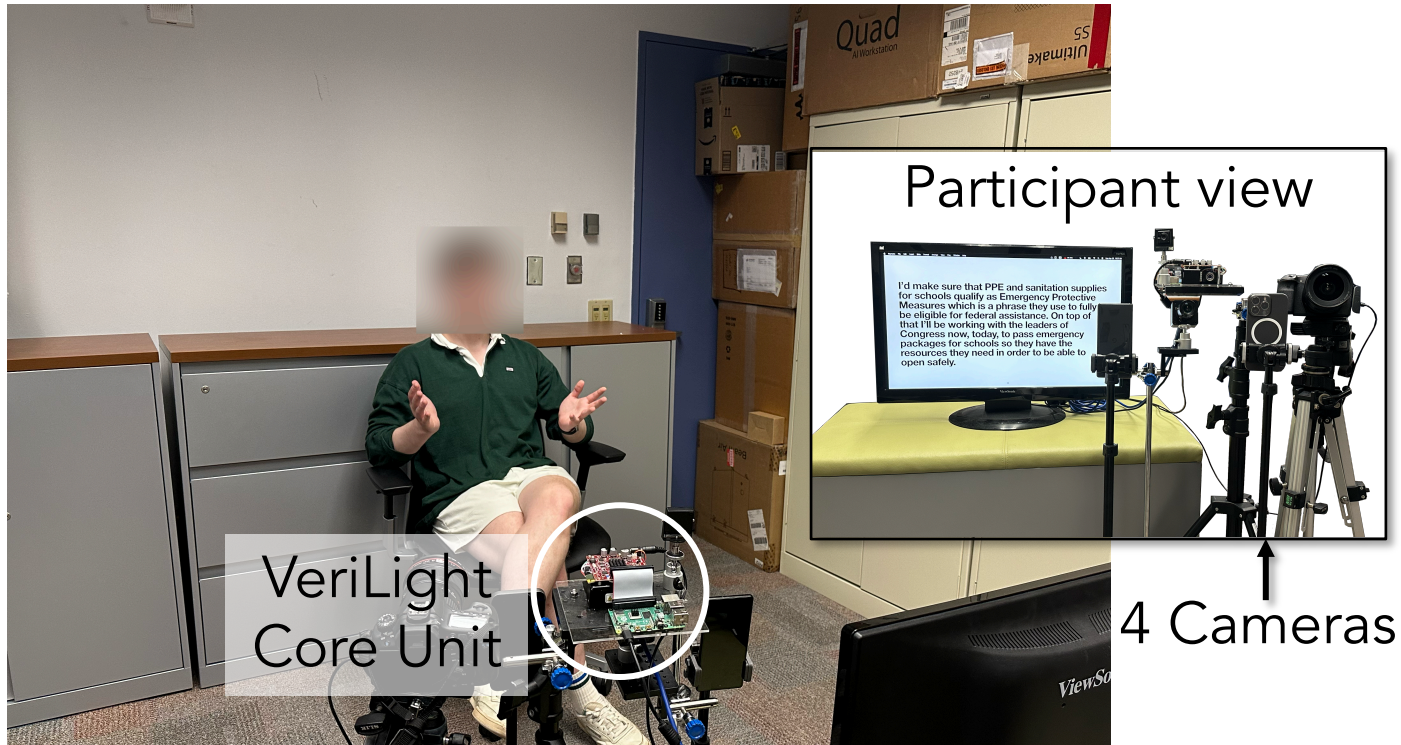


- Error correcting codes
- Closed-loop adaptive embedding algorithm

Evaluation

Protection Performance: Deepfakes

Data



Evaluation

Protection Performance: Deepfakes

Data



- 474 real
- 2,356 deepfakes

- | | | |
|--------------------|---|-------------|
| 1,883 | { | • DaGAN |
| Lip/face
motion | | • SadTalker |
| | | • FOMM |
| | | • TalkLip |
| 473 | { | • FSGAN |
| ID swap | | |

Evaluation

Protection Performance: Deepfakes

Data

- 474 real
- 2,356 deepfakes
 - DaGAN
 - SadTalker
 - FOMM
 - TalkLip
 - FSGAN

Metrics

- Area Under Curve (AUC)
- True Positive Rate (TPR)

Deepfake = Positive

Evaluation

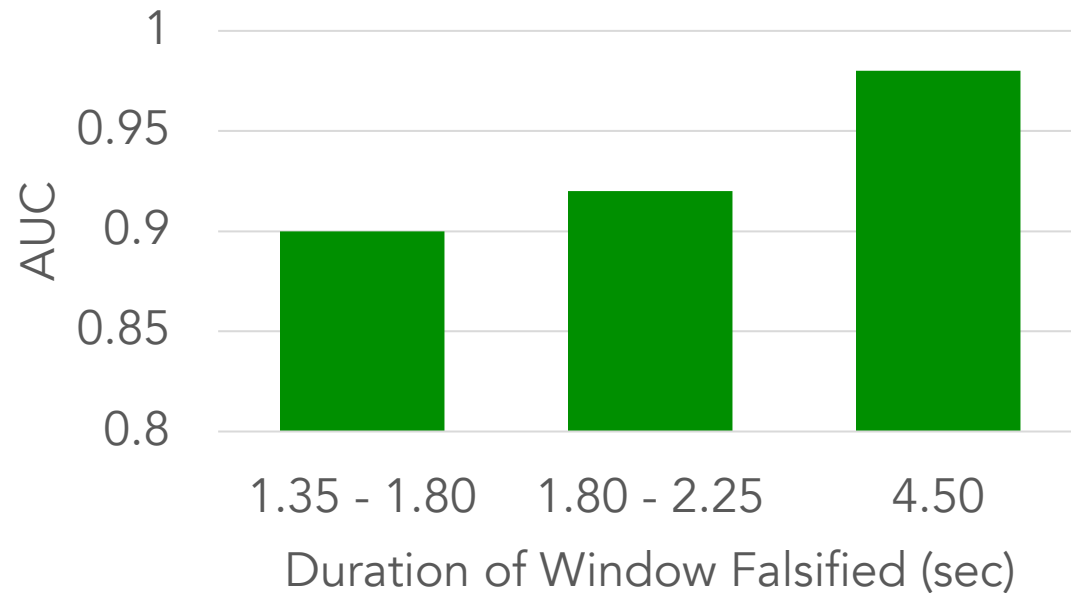
Protection Performance: Deepfakes

Deepfake Model	AUC
DaGAN	0.99
SadTalker	0.99
FOMM	0.99
TalkLip	0.99
FSGAN	1.00

- ✓ AUCs ≥ 0.99
- ✓ 100% TPR
- ✓ 37% average improvement over 10 SOTA passive detectors

Evaluation

Protection Performance: Robustness



✓ AUCs ≥ 0.90

Evaluation

Signature Embedding Robustness

Error-free data extraction:

- ✓ 7 projection surfaces
 - ✓ 5 lighting conditions
 - ✓ 5 recording devices
 - ✓ Up to 5 m, 60°
 - ✓ Transcoding/compression
 - ✓ Aesthetic filters
- and more...



Sample Projection Environments

Evaluation

Signature Embedding Robustness

Error-free data extraction:

- ✓ 7 projection surfaces
 - ✓ 5 lighting conditions
 - ✓ 5 recording devices
 - ✓ Up to 5 m, 60°
 - ✓ Transcoding/compression
 - ✓ Aesthetic filters
- and more...

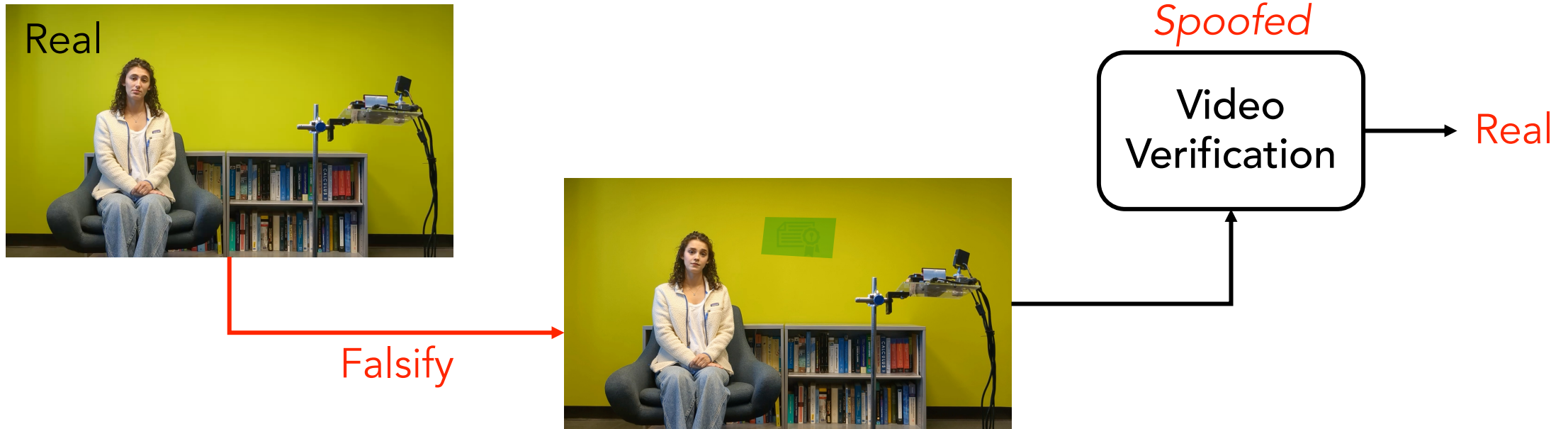


Resilient against various
post-processing techniques
+ environment and
recording conditions

Sample Projection Environments

Evaluation

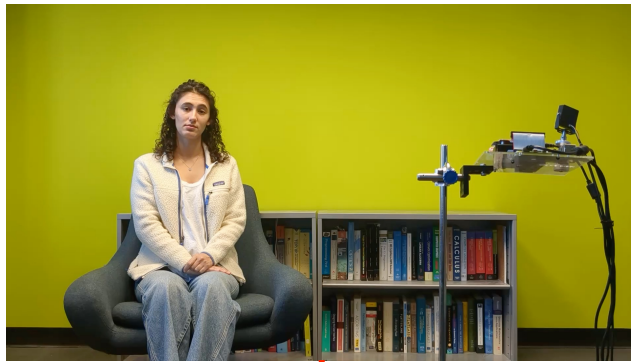
White-Box Adversarial Attacks



Adversarial example: Content visually different
But video descriptor features similar to those of real video

Evaluation

White-Box Adversarial Attacks



Spoofed



Real

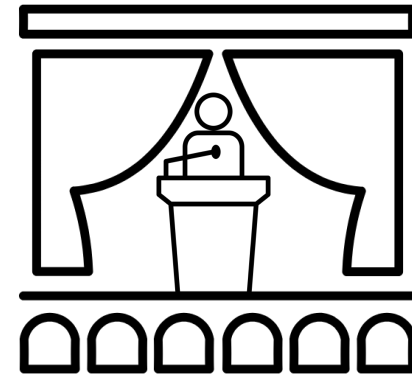
✓ 0% success rates

Future Work

Acoustic embedding
and audio features



Beyond lip/facial
motion and identity



Thank You!

Code, demo, and more on:

<https://mobilex.cs.columbia.edu/verilight/>



Combating Falsification of Speech Videos with Live Optical Signatures

Hadleigh Schwartz¹, Xiaofeng Yan¹, Charles J. Carver², Xia Zhou¹

¹Columbia University ²Massachusetts Institute of Technology
ACM CCS 2025

[Paper](#) [arXiv](#) [Code](#)



VeriLight is a platform that protects high-profile speech videos from visual falsification. It creates signatures physically at live speech events using imperceptible modulated light, so that they are naturally embedded into all authentic recordings.