



CALIFORNIA
Fair Political Practices Commission

April 2025

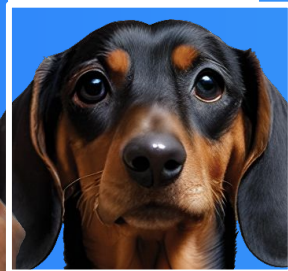
The AI “Antivirus” for Generated Content

Kevin Guo
CEO & Co-Founder, Hive

Not AI-Generated



AI-Generated





AI solutions to understand, search & generate content

Content Moderation

Visual Moderation

Text Moderation

LLM-Based Moderation

Audio Moderation

CSAM Detection

Content Authenticity

AI-Generated Image & Video Detection

Deepfake Image & Video Detection

AI-Generated Audio Detection

AI-Generated Music Detection

AI-Generated Text Detection

IP & Publicity Detection

Media Search

Likeness Detection

Celebrity Recognition

Applications

Moderation Dashboard
End-to-end moderation review tool

AutoML
Custom AI model training

Trusted By:



Detecting AI generated content...



2021



2022



2023

... it used to be obvious...

Detecting AI generated content...



2024



2024



2024

... it used to be obvious... not anymore

Today, it is increasingly difficult to distinguish with the human eye alone



So live today so that future today will past today, as it is tomorrow.



September 2024

 **Gavin Newsom** 
@GavinNewsom · Follow

Manipulating a voice in an "ad" like this one should be illegal.

I'll be signing a bill in a matter of weeks to make sure it is.





Elon Musk retweets altered Kamala Harris campaign ad

In the video, Harris seemingly exposes herself as an incompetent candidate for president. The origin of the video isn't known yet.

8:47 PM · Jul 26, 2024


 128K  Reply  Copy link

[Read 21.6K replies](#)

 **Governor Gavin Newsom** 
@CAGovernor · Follow

California is standing up to harmful deepfakes in political ads and other online content.

To help safeguard the integrity of our elections, these three new laws will remove deceptive content from large online platforms, increase accountability, and better inform voters.



0:25 / 0:39

April 2025

NO FAKES Act

Blackburn, Coons, Salazar, Dean, Colleagues Introduce “NO FAKES Act” to Protect Individuals and Creators from Digital Replicas

April 9, 2025

ENDORSEMENTS

This legislation is endorsed by the Recording Industry Association of America; Motion Picture Association; SAG-AFTRA; YouTube; Recording Academy; OpenAI; Warner Music Group; Universal Music Group; Sony Music; The Walt Disney Company; IBM; Vermillion; **Hive**; Independent Film & Television Alliance; WME; Creative Artists Agency; Human Artistry Campaign; National Association of Broadcasters; the Model Alliance; ASCAP; Nashville Songwriters Association International; the Authors Guild; the National Center on Sexual Exploitation; Television Academy; Enough is Enough; American Association of Independent Music; and more.

stealing their voice and likeness,” said Senator Coons. “Incredible technology like AI can help us push the limits of human creativity, but only if we protect Americans from those who would use it to harm our communities. I am grateful for the bipartisan partnership of Senators Blackburn, Klobuchar, and Tillis, the support of colleagues in the House, and the endorsements of leaders in the entertainment industry, the labor community, and firms at the cutting edge of AI technology.”

“While AI presents extraordinary opportunities for technological advancement, it also poses some new problems, including the unauthorized replication of the voice and visual likeness of individuals, such as artists,” said Senator Tillis. “We must protect against such misuse, and I’m proud to co-introduce this bipartisan legislation to create safeguards from AI, which will result in greater protections for individuals and that which defines them.”

“Americans from all walks of life are increasingly seeing AI being used to create deepfakes in ads, images, music,

TAKE IT DOWN Act

The Honorable Mike Johnson
Speaker
US House of Representatives
Washington, DC 20515

The Honorable Hakeem Jeffries
Minority Leader
US House of Representatives
Washington, DC 20515

Dear Speaker Johnson, Leader Scalise, Leader Jeffries

We, the undersigned technology companies and trade support for the TAKE IT DOWN Act. We value the leadership of Speaker Johnson, Leader Scalise, and Leader Jeffries, and Representatives Salazar and Dean for their removal of non-consensual intimate images from online platforms and criminalizes the publication of such images, when

Tech companies see firsthand the extraordinary benefits of AI. Unfortunately, as with any technology, bad actors are exploiting it - including by using AI to create non-consensual intimate images of people. Passage of this legislation will help protect our privacy and hold bad actors accountable, and it will better equip

Urgent action on this legislation would echo calls from President Trump and First Lady Melania Trump's Executive Order. I have followed the Senate's lead by passing the Commerce Committee. Now, it is your turn to pass the bill for President Trump's signature. We look forward to the House

Sincerely,
Amazon
Entertainment Software Association
Google
Hive
Meta
Software & Information Industry Association
TechNet
X Corp

The Honorable Steve Scalise
Majority Leader
US House of Representatives
Washington, DC 20515

The Honorable Katherine Clark
Minority Leader
US House of Representatives
Washington, DC 20515

Sincerely,

Amazon
Entertainment Software Association
Google
Hive
Meta
Software & Information Industry Association
TechNet
X Corp

AI “antivirus” to detect the misuse of AI and Deepfakes



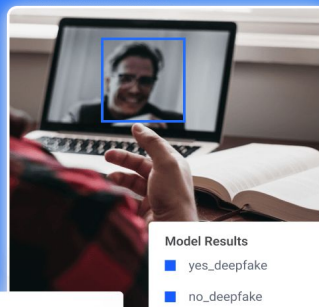
AI-Generated Image & Video Detection

Model Results	
■ ai_generated	1.00
■ flux	1.00



Model Results	
■ ai_generated	0.96
■ flux	0.49

Deepfake Image & Video Detection

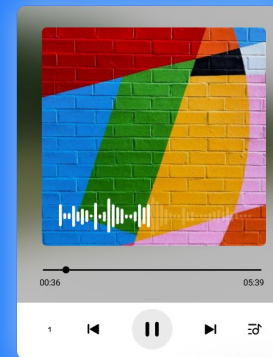


Model Results	
■ yes_deepfake	0.942
■ no_deepfake	0.000



Model Results	
■ ai_generated	0.96

Synthetic Audio & Music Detection



Model Results	
■ ai_generated	0.932
■ not_ai_generated	0.864

AI “antivirus” to detect the misuse of AI and Deepfakes



Organic or Diffused: Can We Distinguish Human Art from AI-generated Images?

Anna Yoo Jeong Ha*, Josephine Passananti[†], Ronik Bhaskar, Shawn Shan, Reid Southern¹, Haitao Zheng, Ben Y. Zhao
Department of Computer Science, University of Chicago,¹ Concept Artist
{annah, josephinep, rbhaskar, shawnshan, htzheng, ravenben}@cs.uchicago.edu,
¹ReidSouthern@gmail.com

- Supervised classification does surprisingly well, and Hive outperforms all detectors (human and ML), and produces zero false positives. Unsurprisingly, accuracy seems to correlate with expected training data availability: biggest classifier (Hive) performs best; all classifiers perform the worst on Firefly, the newest of the major generative models.

Organic or Diffused: Can We Distinguish Human Art from AI-generated Images?

Tested on Human Artworks + AI-generated Images			
Detector	ACC (%) ↑	FPR (%) ↓	FNR (%) ↓
Hive	98.03	0.00	3.17
Optic	90.67	24.47	1.15
Illuminary	72.65	67.40	4.69
DE-FAKE	50.32	41.79	56.00
DIRE (a)	55.40	99.29	0.86
DIRE (b)	51.59	25.36	66.86
Ensemble	98.75	0.48	1.71

Table 1: Performance of automated detectors tested on unperturbed human artworks and AI-generated images. The Ensemble detector (Hive+Optic+Illuminary) takes scores from Hive, Optic, and Illuminary, using the highest confidence value as the determining score.

“Hive is the clear winner”

A 2024 independent research study from the University of Chicago found that our AI Detection model outperforms against competing models and human expert analysis

The US Department of Defense is investing in deepfake detection

It has signed a contract with startup Hive AI worth \$2.4 million over two years.

By Melissa Heikkilä

December 5, 2024



STEPHANIE ARNETT/MIT TECHNOLOGY REVIEW

- Awarded DoD contract from 35 competing solutions
- Supporting DIU for:
 - AI-Generated Image & Video Detection
 - AI-Generated Audio Detection
 - Deepfake Image & Video Detection
 - Liveness Detection

Singapore General Elections

Rise of AI and deepfakes in lead-up to GE2025

Digitally manipulated videos and photos of politicians will be outlawed during the upcoming elections period. Meanwhile, they are already proliferating and blurring the lines between fact and fiction.

Published Apr 10, 2025

SDP contesting Marsiling-Yew Tee

AI GENERATED

When you get

AI GENERATED

AI GENERATED

Opposition MPs react to negative remarks from PAP

AI GENERATED

and I feel that something is going wrong

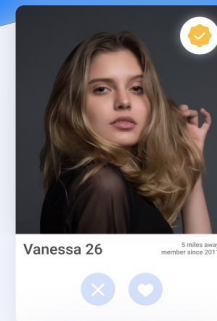
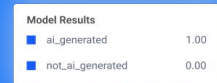
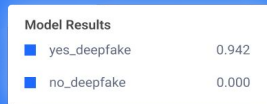
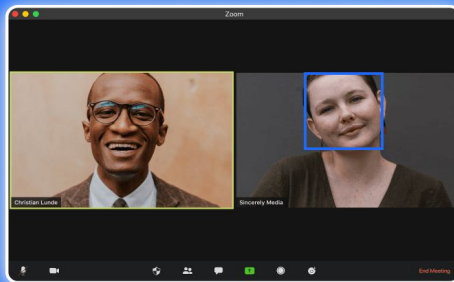
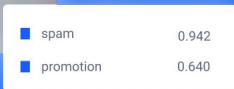
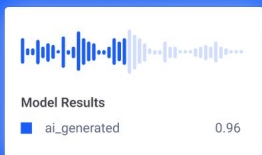
This TikTok video of Singapore's late founding prime minister Lee Kuan Yew, for example, has received 176,700 views as of the afternoon of Apr 9.

AI GENERATED

Opposition MPs react to negative remarks from PAP

Opposition MPs react to negative remarks from PAP

It's becoming harder to tell what is real...



... new tools are required to detect the “imperceptible”

Q&A