A.I. Deepfakes And Chatbots

Deepfakes and Elections

What is a deepfake and how are they used in an elections context?

MORGAN MEAKER BUSINESS 03.10.2023 12:00 PM

Slovakia's Election Deepfakes Show AI Is a Danger to Democracy

Fact-checkers scrambled to deal with faked audio recordings released days before a tight election, in a warning for other countries with looming votes.

Progressive Slovakia party leader Michal Simecka. PHOTOGRAPH: ZUZANA GOGOVA/GETTY IMAGES

COMMITMENT 2024

"DEEPFAKE" ROBOCALLER

NH ATTORNEY GENERAL

- Someone using President Biden's voice
- Says "your vote makes a difference in November, not this Tuesday"
- AG telling people to disregard message and contact them

JOE BIDEN

5 abc WCVB

YOUR LOCAL ELECTION HQ

North Carolina 6th District candidate Mark Walker calls video shared by PAC a 'deepfake'

by: Emily Mikkelsen
Posted: Feb 29, 2024 / 11:30 AM EST
Updated: Mar 1, 2024 / 06:45 AM EST

Long Term Consequences

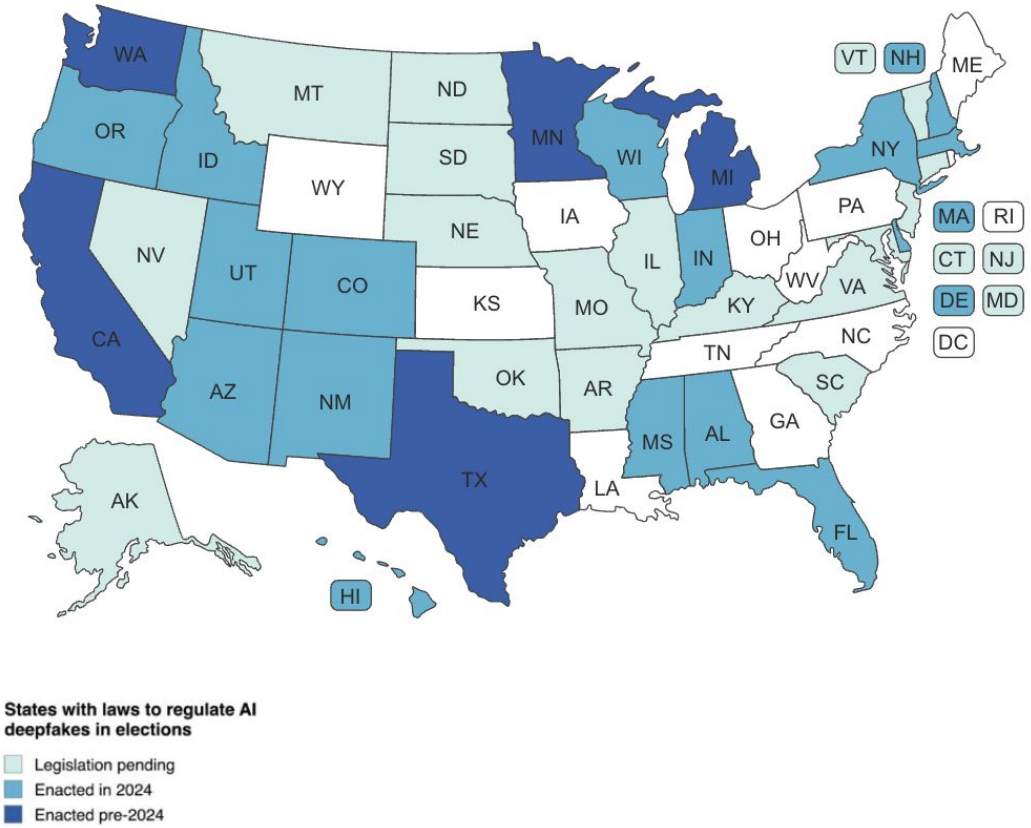
Urgency

- Rapidly improving in quality and quantity
- Accessibility
- You only need one bad elections deepfake

Legislation Addressing this Issue

- 21 States have passed legislation
- 49 States and D.C. have introduced legislation
  - Mostly in this session and last
  - Most that did not pass was because of time
- 18 States have already introduced this legislation this session
- Strong bipartisan support, unanimous votes

States where election deepfake bills passed or are in play as of Jan. 2025...



https://www.citizen.org/article/tracker-legislation-on-deepfakes-in-elections/

Key Elements of Anti-Fraudulent Deepfake Legislation

- Prohibit distribution of unlabeled deepfakes - **Why disclosure not ban?**
- Standards for disclosure - must be very clear and prominent
- Covers all people – not just candidates, parties and committees e.g. influencers
- Usually within a certain number of days of E-day
- Only covers people who knowingly circulate a deepfake
- Establish a right for affected parties to seek injunction to take it down
- Establishes penalties

Protections

- No liability for broadcasters or platforms that make reasonable effort to prevent deepfakes, or that show deepfakes as part of news coverage and describe as deepfakes
- Satire and parody are protected
- Individuals who are unknowingly reposting are not held liable

Lessons Learned

- **Clear** definition is critical
  - Depicts someone doing or saying something they never did or said
  - Provides a fundamentally different understanding of the person's speech or behavior
  - Intent to undermine candidates reputation or otherwise deceive voters

- Satire and parody exemptions
- Liability has to be on the distributor

Non-Consensual Intimate Deepfakes

What is an intimate deepfake?

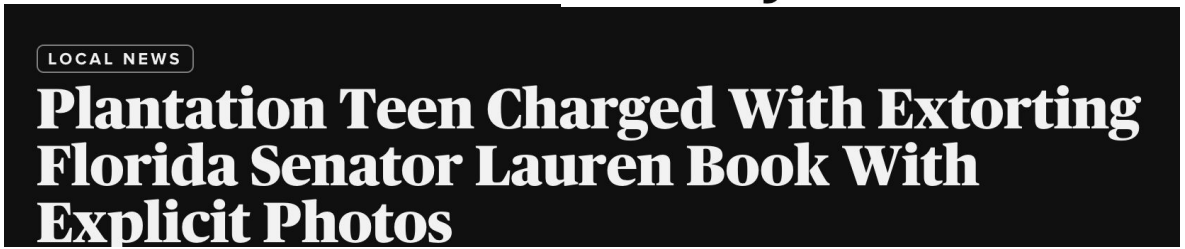
Pervasiveness of this Problem

- Intimate deepfakes make up ~98% of all deepfakes
- ~99% of victims are women
- Children (girls) are a significant percentage of the victims
- A study recently found that 57% of those under 18 years of age are concerned about becoming victims to intimate deepfakes and 10% of individuals reported being a victim of intimate deepfakes, knowing a victim, or both.
- WIRED found on Telegram at least 50 bots that claim to create explicit photos or videos of people with only a couple of clicks - these bots have over 4 million monthly users

Examples

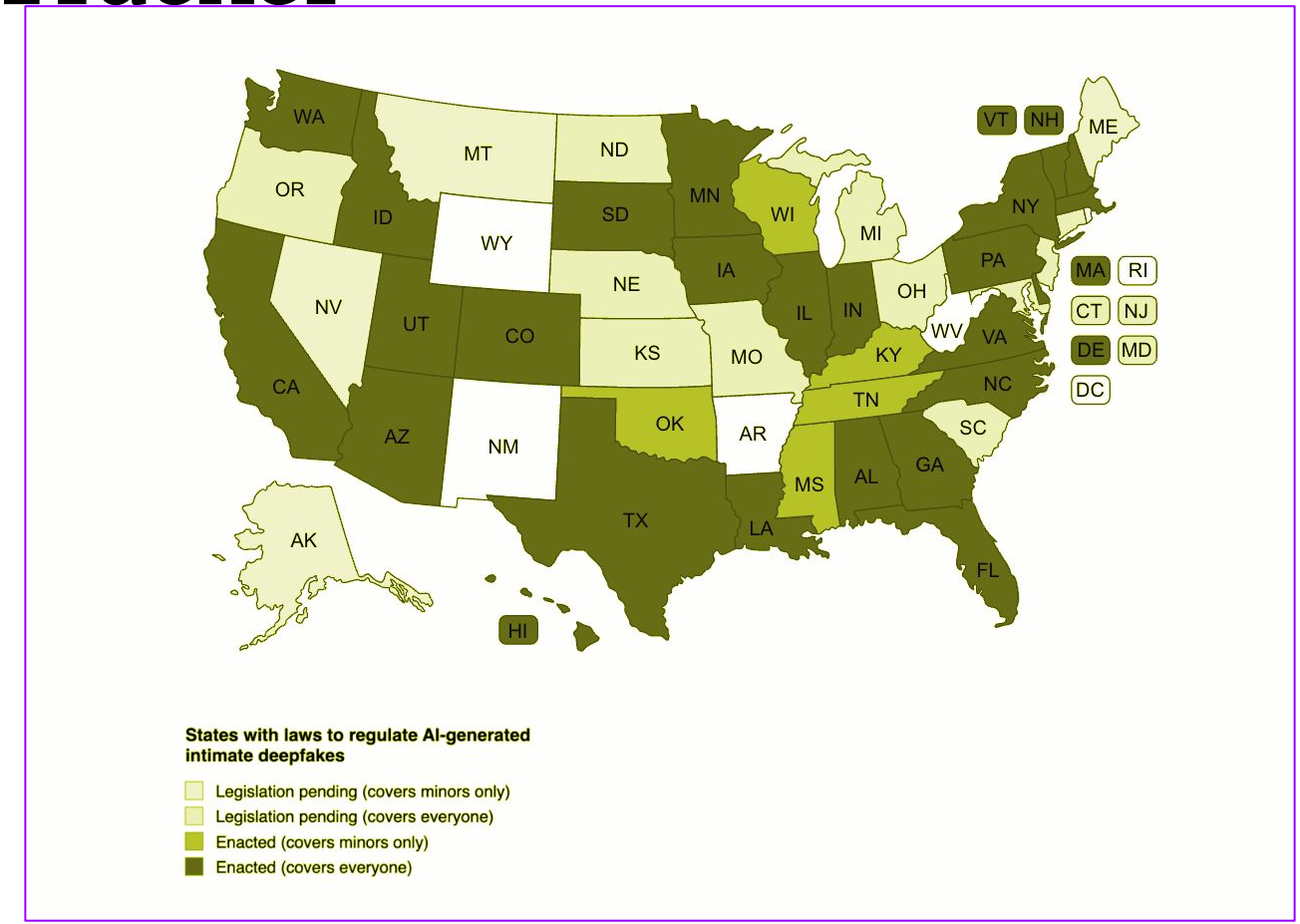
AI-generated fake nude photos of girls from Winnipeg school posted online

South Korea has jailed a man for using AI to create sexual images of children in a first for country's courts




harms
contempt
depression
disconnected
disrespected
PTSD
job loss
isolated
withdrawn
panic
silenced
dissociation
mistrustful
anxiety

State Tracker



Legislation Passed in the States

- 30 states have enacted legislation + 1 more just passed
- 45 states have introduced legislation
- Broad bipartisan support

Model Legislation

A few important recommendations to keep in mind when drafting this legislation:

1. Provide for both civil liability and criminal penalties. Civil liability can afford the victim injunctive and economic relief while the criminal penalty can act as a deterrent.
2. Allow for a defense of consent but require more than mere assertions of oral consent to satisfy it.
3. Disclaimers of inauthenticity or unauthorized creation should not be permitted as a defense against intimate deepfakes,as such disclosure does not mitigate the reputational and psychological harm done to the victims.

https://www.citizen.org/article/public-citizen-model-state-law-regulating-non-consensual-intimate-deepfakes/

Other Legislative Approaches

Consumer Protection + Chatbot Labeling

- Human-seeming chatbots, “agents” and avatars are widespread and poised to become pervasive.
- It is often hard or impossible for consumers to know when they are dealing with a human or computer.
- AI has the ability to manipulate people, especially when they’re guard is down because they believe they are communicating with a person.
- Model legislation establishes that it is an unfair and deceptive practice not to disclose to consumers when they are engaging with a human-seeming AI.

Questions?

Ilana Beller:
ibeller@citizen.org