



**CENTRE FOR INNOVATION**  
Leiden University



**Universiteit Leiden**



# Week 5: The Art of Hacking: Of Deepfakes and Toastercats

Jeanine Reutemann

## Inception v3, trained on ImageNet

Enter a valid image URL or select an image from the dropdown:

enter image url

<http://i.imgur.com/il0yXAA.png>

or select image

Use GPU

Show computation flow



toaster | 98%  
Crock Pot | 1%  
Siamese cat | 0%  
wallaby | 0%  
carton | 0%

**Deepfakes;**

The background is a low-poly geometric pattern composed of numerous triangles. The color palette is a gradient: light blue and white at the top, transitioning through pink and purple to dark purple and black at the bottom. The text 'Deepfakes;' is positioned in the upper left quadrant.



**ORIGINAL**

**DERPFAKES**

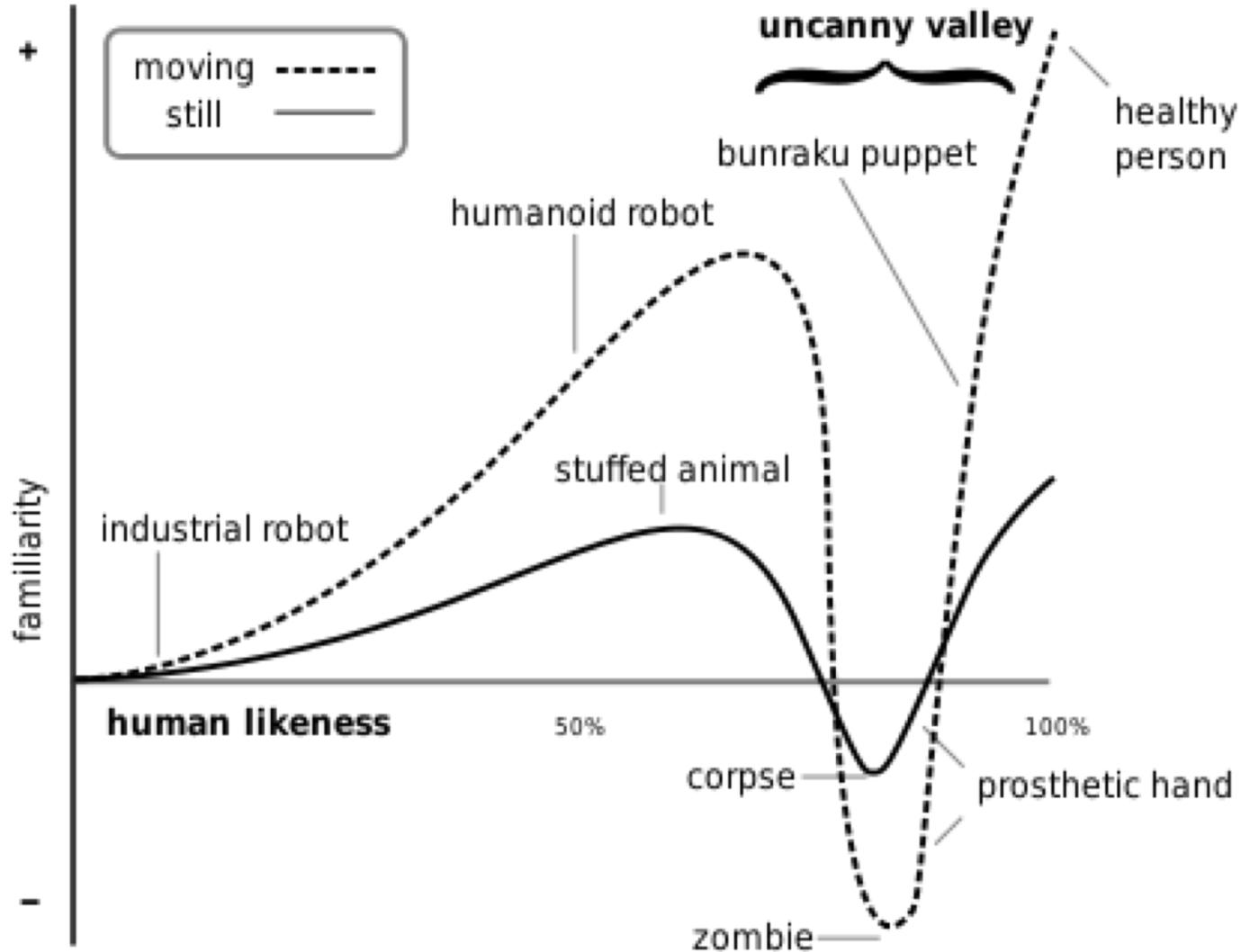


★ **THE PRESIDENT**

★ **THE PRESIDENT**



# Deepfakes: Uncanny Valley Effect

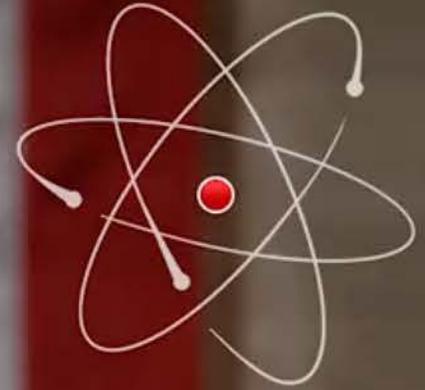


# CGI (Computer Generated Imagery)



The background is a complex, low-poly geometric pattern. It features a color gradient that starts with light blue and white at the top left, transitions through various shades of purple and magenta, and ends in dark red and black at the bottom right. The shapes are irregular polygons of varying sizes, creating a textured, crystalline appearance.

**Deepfakes;  
Audio-to-Lipsync;**



# TWO MINUTE PAPERS

WITH KÁROLY ZSOLNAI-FEHÉR (KZF)

## **AUDIO TO OBAMA: LEARNING LIP SYNC FROM AUDIO**

Disclaimer: I was not part of this research project, I am merely providing commentary on this work.



**Deepfakes;  
Audio-to-Lipsync;  
Face-to-Face;**

# Face2Face: Real-time Face Capture and Reenactment of RGB Videos

*Justus Thies<sup>1</sup>, Michael Zollhöfer<sup>2</sup>,  
Marc Stamminger<sup>1</sup>, Christian Theobalt<sup>2</sup>,  
Matthias Nießner<sup>3</sup>*

<sup>1</sup>University of Erlangen-Nuremberg

<sup>2</sup>Max-Planck-Institute for Informatics

<sup>3</sup>Stanford University

**CVPR 2016 (Oral)**



**Deepfakes;  
Audio-to-Lipsync;  
Face-to-Face;  
Text-to-Speech;**





**Deepfakes;  
Audio-to-Lipsync;  
Face-to-Face;  
Text-to-Speech;  
Facial Recognition;**

**FaceApp  
Snapchat  
Etc.**



**Facial-Recognition meets  
Facial Expressions meets  
Emotion Recognition**



**Deepfakes;  
Audio-to-Lipsync;  
Face-to-Face;  
Text-to-Speech;  
Facial Recognition;  
Digital Nudging;**

# Push-Notification Red-Colour Symbols Nudging via: Interaction Design



**“[...] the use of user interface design elements to guide people's choices or influence users' inputs in online decision environments”**

**Weinman, Schneider and Brocke, 2015**

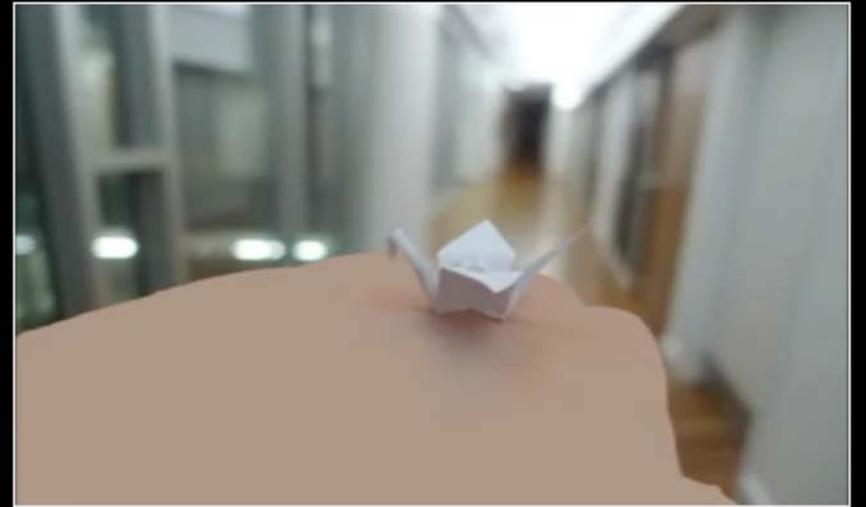


**Deepfakes;  
Audio-to-Lipsync;  
Face-to-Face;  
Text-to-Speech;  
Facial Recognition;  
Digital Nudging;  
[...]**

**Next things to come (October 2018)?**



original video of a day with light breeze



original origami bird video

# Adobe Cloak is Content-Aware Fill for Video

**In a few months from now,  
we won't be able to distinguish  
between deepfakes and 'normal'  
videos.**

**Are we?**

**YES & NO**

# Continuous Race

**«Then, one afternoon, while studying several deepfakes, Lyu realized that the faces made using deepfakes rarely, if ever, blink. And when they do blink, the eye-movement is unnatural. This is because deepfakes are trained on still images, which tend to show a person with his or her eyes open.»**

**Siwei Lyu, State University of New York at Albany**



BBC News Africa

@BBCAfrica

Folgen



## THREAD

In July 2018, a horrifying video began to circulate on social media.

2 women & 2 young children are led away by a group of soldiers. They are blindfolded, forced to the ground, and shot 22 times.

**#BBCAfricaEye** investigated this atrocity. This is what we found...

Tweet übersetzen



<https://twitter.com/BBCAfrica/status/1044186344153583616>

# VERIFICATION STRATEGIES?

**«Amnesty International is already grappling with some of these issues. Its Citizen Evidence Lab verifies videos and images of alleged human-rights abuses. It uses Google Earth to examine background landscapes and to test whether a video or image was captured when and where it claims. It uses Wolfram Alpha, a search engine, to cross-reference historical weather conditions against those claimed in the video.»**

**Fake news: you ain't seen nothing yet,**

<https://www.economist.com/news/science-and-technology/21724370-generating-convincing-audio-and-video-fake-events-fake-news-you-aint-seen>

# **VERIFICATION STRATEGIES?**

**Media: unique key that only the signing organisation—or the originating device—possesses.**

# **VERIFICATION STRATEGIES?**

**Decentralized timestamping on the blockchain**

# VERIFICATION STRATEGIES?

**«In an initial round of testing last June, researchers were able to identify “speaker inconsistencies and scene inconsistencies,” two markers of video that’s been tampered with, with 75% accuracy in a set of hundreds of test videos.»**

**«DARPA is funding new tech that can identify manipulated videos and ‘deepfakes’» 1th of May 18,  
<https://techcrunch.com/2018/04/30/deepfakes-fake-videos-darpa-sri-international-media-forensics/>**

**All of these idea  
solutions will have  
no impact on our  
fakenews issue.**



AMERICA TOUR

TRUTH LIBRARY .ORG

# Inception v3, trained on ImageNet

Enter a valid image URL or select an image from the dropdown:

enter image url

<http://i.imgur.com/il0yXAA.png>

or select image ▼



Use GPU



Show computation flow



toaster | 98%

Crock Pot | 1%

Siamese cat | 0%

wallaby | 0%

carton | 0%

**Medium.com: Machine Learning is Fun Part 8:  
How to Intentionally Trick Neural Networks  
A Look into the Future of Hacking**