AI evolution and impact on security

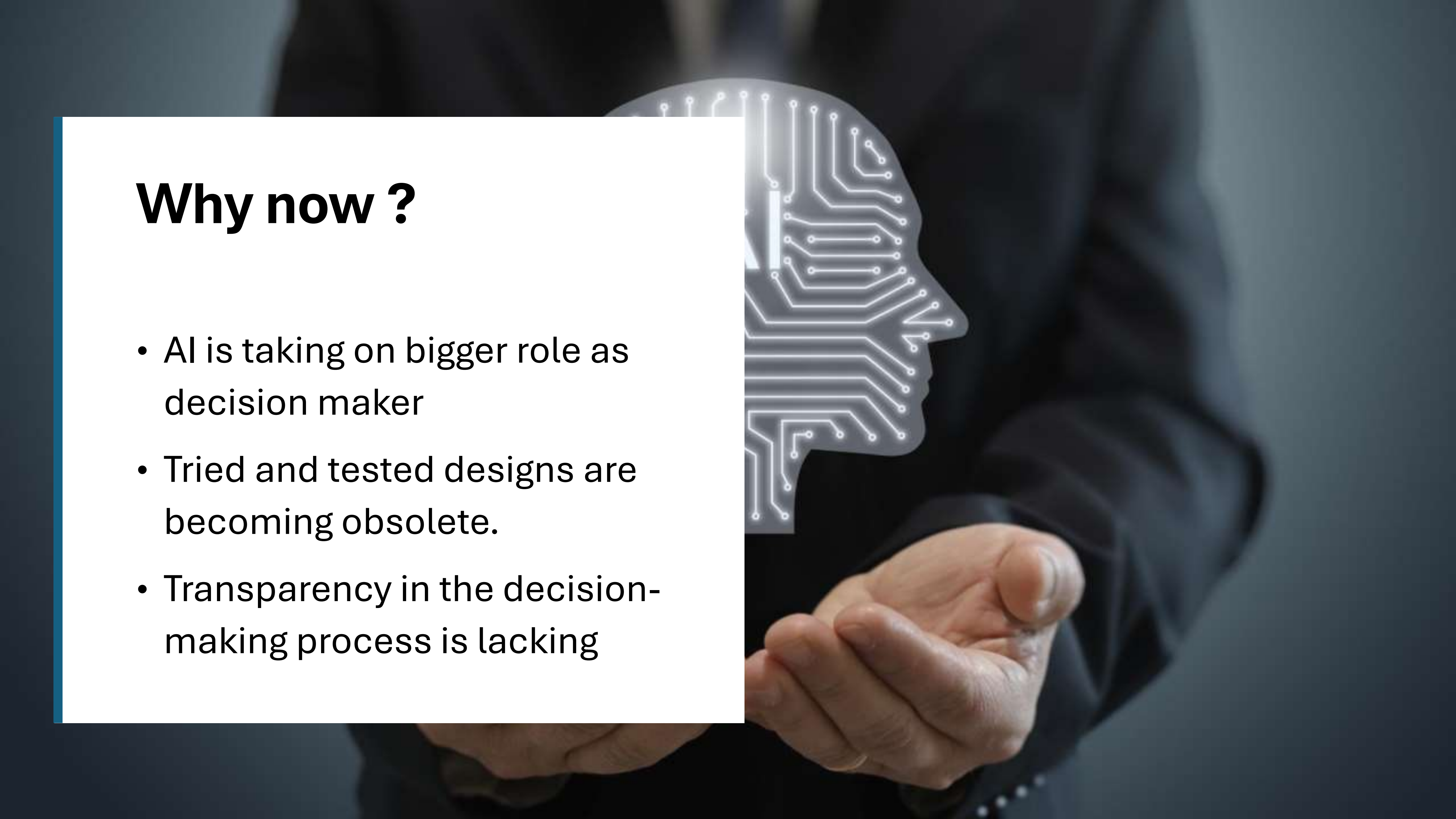
Arnar S. Gunnarsson

Preface

- This talk might sound negative and against AI
- I use most of the main tools myself
    - ChatGPT
    - CoPilot
    - Eleven Labs
    - more

Why now ?

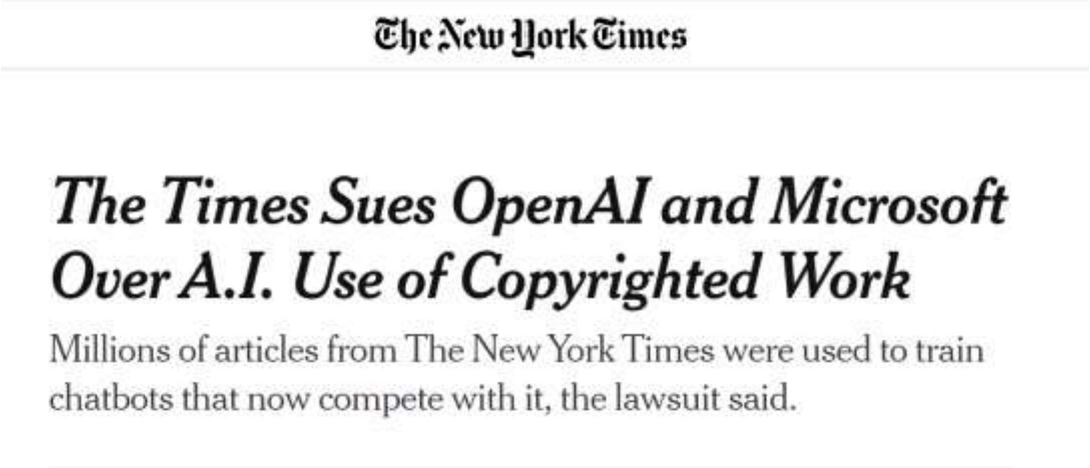
- AI is taking on bigger role as decision maker
- Tried and tested designs are becoming obsolete.
- Transparency in the decision-making process is lacking



AI and Security

Accountability and traceability

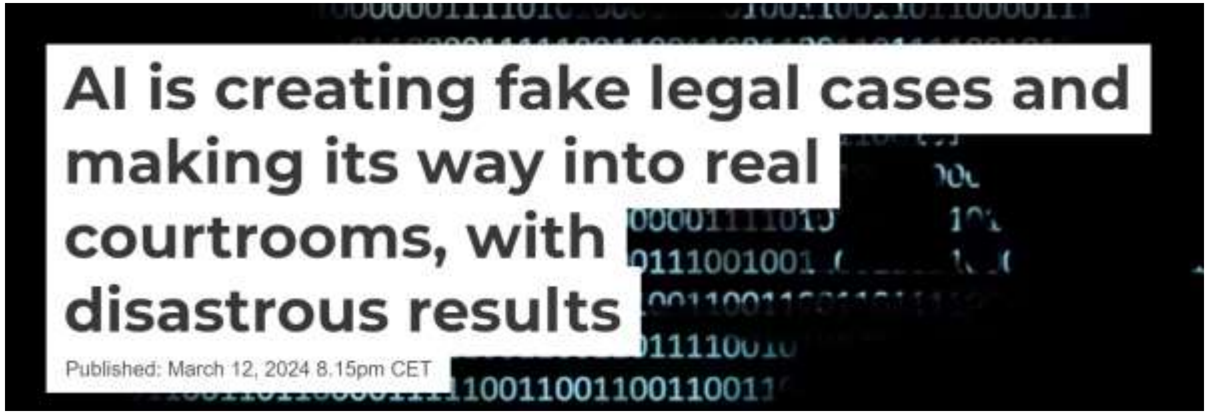
- Black box models
- What happens inside ?
- How was the decision made ?
- Is the model discriminating certain demographic groups
- Developer decision biases ?
- Malicious intent ?



The New York Times

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said.

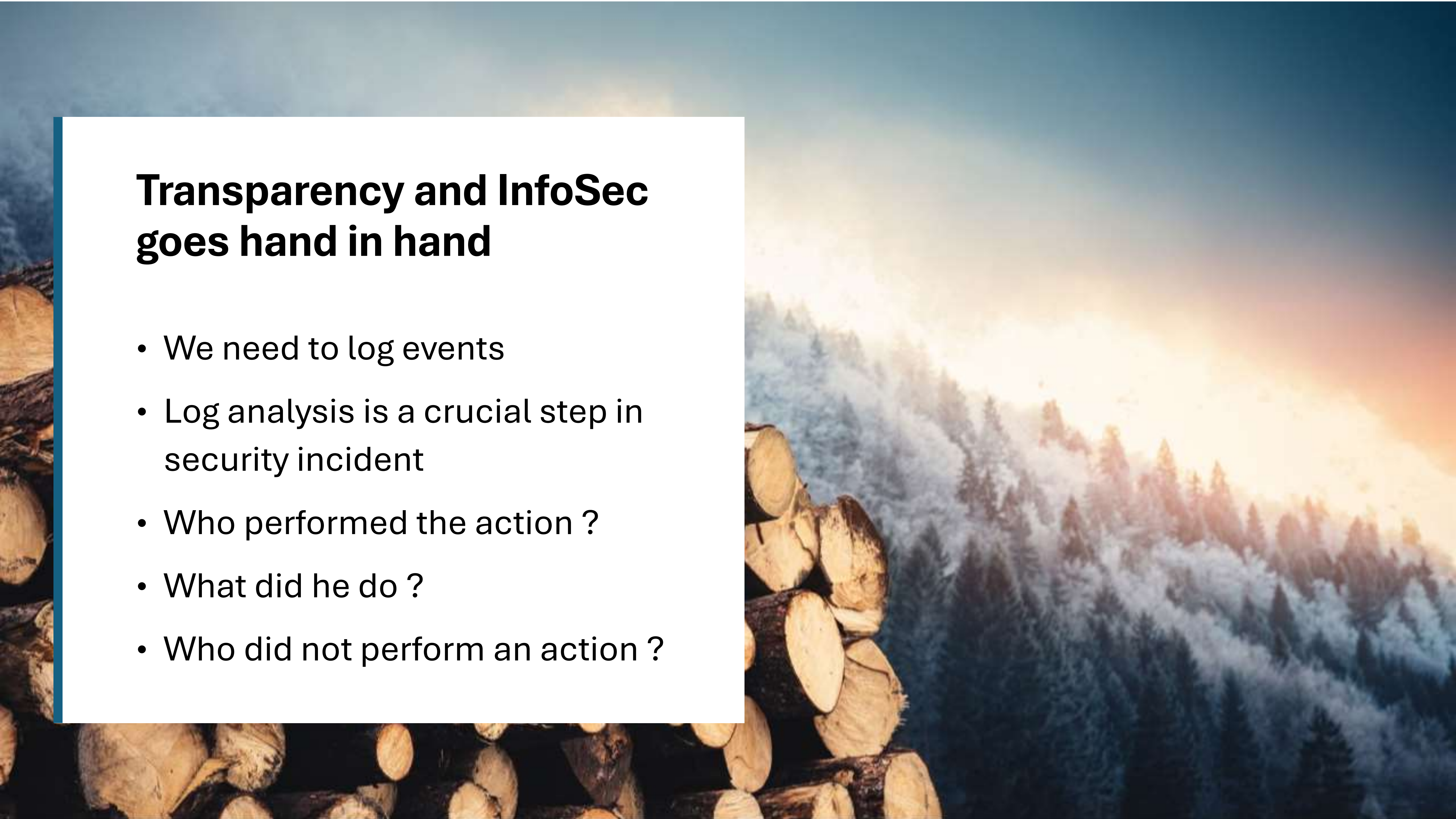


AI is creating fake legal cases and making its way into real courtrooms, with disastrous results

Published: March 12, 2024 8.15pm CET

AI in security

- Security teams need AI
    - Chronic alert fatigue is real
    - Most alerts are false positive (80%)
- Security and Ops teams are always busy
    - Some AI based security tools learn bad behaviour
- Controlant has been using AI to analyze all logs for security indicators since May 2022
    - 45.000.000 events per hour



Transparency and InfoSec goes hand in hand

- We need to log events
- Log analysis is a crucial step in security incident
- Who performed the action ?
- What did he do ?
- Who did not perform an action ?

Trust and AI

Do we trust AI

“Three in five (61 percent) are wary about trusting AI systems. 67 percent report low to moderate acceptance of AI.”

Source: KPMG, 2023 Global study on the shifting public perceptions of AI

Inherent trust

- Foundational in IT and Security
- Insider Threats
- AI is making obsolete

Zero Trust Access

- Assume compromise
- Verify explicitly
- Least privilege

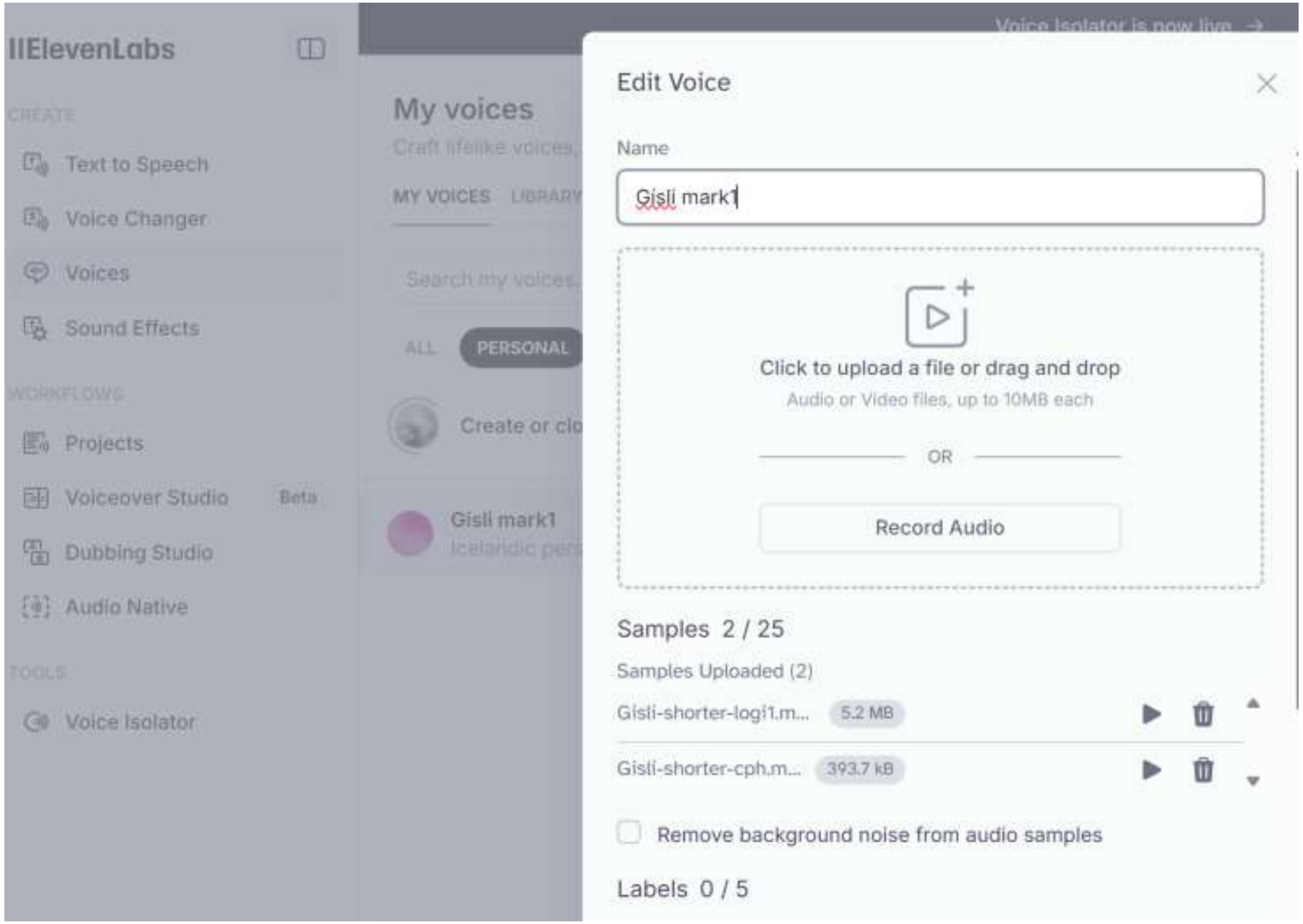
Example – Phishing

SPEAR PHISHING – BANK TRANSFERS

OUR SOLUTION: VOICE RECOGNITION

IF YOU GET A CALL AND YOU KNOW THE VOICE, YOU ASSUME TRUST AND PERFORM THE REQUEST

Example – Eleven Labs



Edit Voice

Name: Gisli mark1

Click to upload a file or drag and drop

Audio or Video files, up to 10MB each

OR

Samples 2 / 25

Samples Uploaded (2)

Gisli-shorter-logi1.m...	5.2 MB
Gisli-shorter-cph.m...	393.7 kB

□ Remove background noise from audio samples

Labels 0 / 5

Complex problem ?
simple solution ?

- Implement manual MFA on the human factor
- Create a list of rotating “passphrase of the week” for authenticating such requests
- Call back and ask, “what is this weeks passphrase”

Fixing the black box problem

Explainable AI - XAI

- Decision tree surrogate model
  - Simpler models
  - Train on input / output of the black box models

- XAI LIME
  - Local Interpretable Model-Agnostic Explanations
- XAI “TRUST”
  - Transparency Relying Upon Statistical Theory
  - 98% accuracy

Thank you