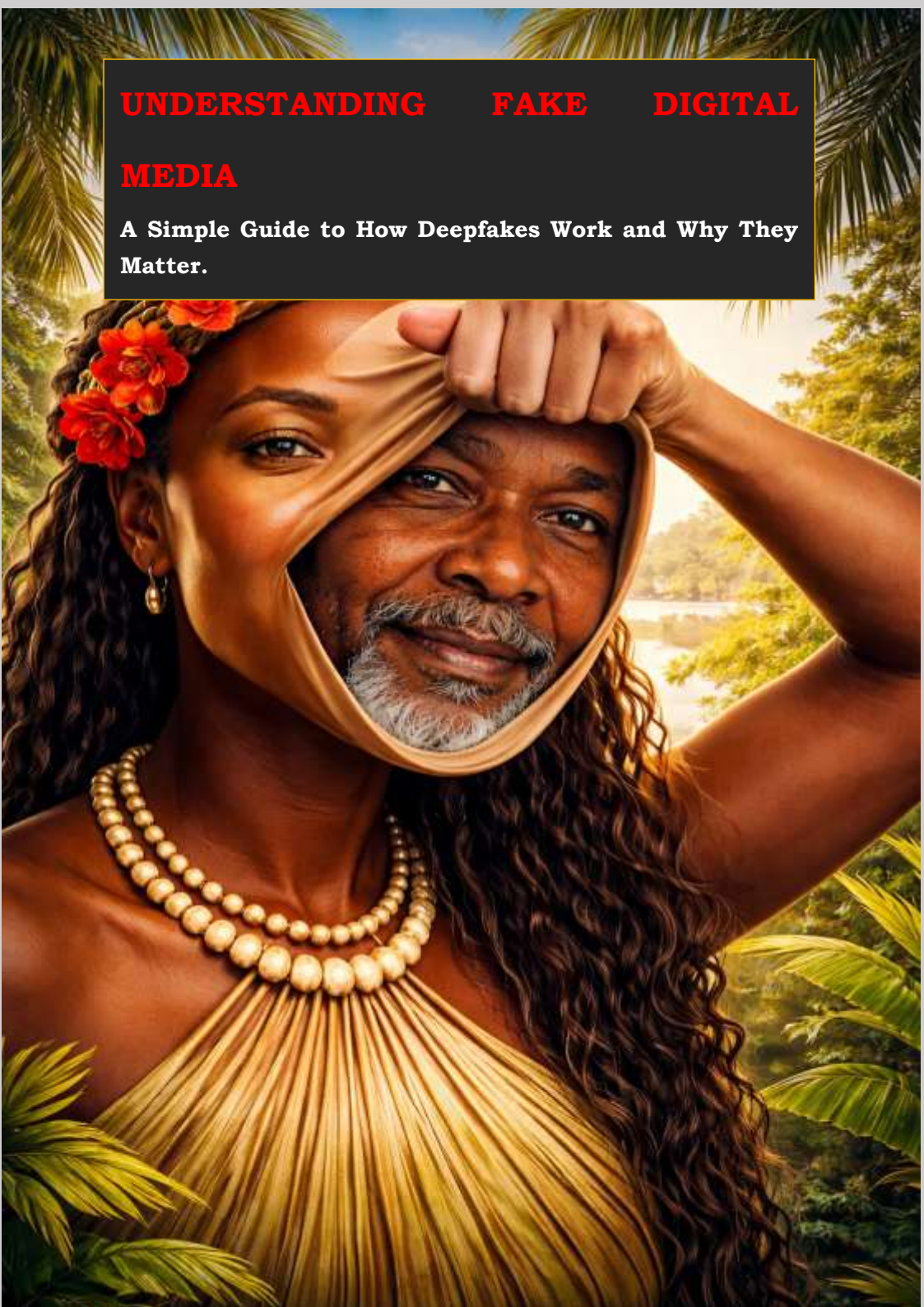

UNDERSTANDING FAKE DIGITAL MEDIA

A Simple Guide to How Deepfakes Work and Why They Matter.

What Are Deepfakes?

A deepfake is a digital photo, video or sound file of a real person that has been edited to create an extremely realistic but false depiction of them doing or saying something that they did not actually do or say. Deepfakes are created using artificial intelligence software that currently draws on a large number of photos or recordings of the person to model and create content.

Deepfake Techniques.

Face Swapping: Replacing a person's face with another in a video or photo using AI.

Voice Cloning: Using AI to mimic someone's voice.

Lip Syncing: Making a person's lips appear to match different audio using AI.



How Deepfakes Are Made?

Deepfakes are created using artificial intelligence (AI) that learns patterns from many images or videos of a person, especially their face. The AI studies these images and learns how the person looks, moves, and speaks.

Two main methods are used:

Generative Adversarial Networks (GANs) – Two AI systems compete with each other. One creates fake images or videos, while the other tries to detect if they are fake. Through repeated practice, the system learns to create very realistic fake content.

Diffusion Models – These models start with a noisy or unclear image and gradually improve it by removing the noise until a realistic image or video is produced. Some can also generate images based on text instructions.

Deepfakes rely on neural networks, which are computer systems designed to recognize patterns. By training on hundreds or thousands of images, the AI learns to recreate a person's face or actions, allowing it to generate realistic but fake photos or videos.

The Dangers of Deepfakes.

While deepfakes can be used for harmless or even educational purposes, they can also be used unethically in the following ways:

Identity theft: Deepfakes can impersonate individuals in videos or audio recordings, potentially allowing criminals to commit fraud under someone else's name.

Threats/blackmail: Deepfakes can create convincing footage or audio that puts the subject in a bad light, which can then be used to threaten or blackmail them.

Cyberbullying: Deepfakes can be used to create embarrassing or hurtful videos or audio recordings of someone.

Propaganda/fake news: Deepfakes can be used to create fake videos or audio recordings that trick people and sway their opinions on important topics.

Election manipulation: Deepfakes can be used to make candidates look bad or confuse voters by creating fake speeches or interviews.

Nonconsensual Intimate Images (NCII): Deepfake photos and videos that remove the subject's clothing without their knowledge or consent.



Where You'll Find Deepfakes. Deepfake videos are most

commonly encountered on:

Platform	% Reporting Deepfakes
Facebook	68%
Instagram	30%
TikTok	28%
X (formerly Twitter)	17%

Source:https://www.mcafee.com/blogs/internet-security/state-of-the-scamiverse/

How to Identify Deepfakes?

Detecting deepfakes involves looking for inconsistencies and anomalies in videos. photos, or audio recordings that may not be immediately apparent. It's important to note that deepfake technology is quickly improving, so **it may soon be impossible to tell a fake from something real.**

.

- **Look closely at people's faces** in the video or picture. If their skin looks too smooth, their eyes don't blink normally, or their facial expressions seem off, it might be a deepfake.

- If the voice doesn't sound quite right, like it's **too robotic or the tone doesn't match how the person usually speaks,** it could be a sign of a deepfake.

- Pay attention to **how the lips move**. If they're not syncing up with the words perfectly, or the movement looks strange, the video might be manipulated.

- Consider the video's **overall appearance**. Sometimes, the area around the person or behind them might look **blurry or glitchy**. This can be a sign that the video has been altered.

- Or if the **shadows on the person's face or in the background don't match up with the lighting** in the rest of the video, it could be a clue that it's a deepfake.

- **Pay attention to the eyes and eyebrows.** Do shadows appear in places that you would expect? DeepFakes often fail to fully represent the natural physics of a scene.

- **Pay attention to facial moles.** Does the mole look real?

- **Pay attention to the size and color of the lips.** Does the size and color match the rest of the person's face?

A variety of **deepfake detectors** are available online,

but as these operate based on a model that calculates probabilities, which **will never be 100% reliable.**

Deepfake and National Security.

Deepfakes pose wide-ranging threats to cybersecurity, national security, and global stability by enabling fraud, disinformation, cyber warfare, and manipulation of evidence, which can disrupt critical infrastructure, damage diplomatic relations, and even trigger conflict. They can also be exploited by terrorists and malicious actors to spread propaganda or coordinate attacks. In Vanuatu we have seen across the media especially Facebook a very rapid increase in the use of deepfakes especially towards leaders both in government, society and churches and it will likely increase in the future.

Implications for Democracy

Deepfakes can spread misinformation and propaganda, causing public confusion, distrust, and manipulation of opinions, including interference in elections and political processes. They can target government officials, influence public perception, and even fabricate evidence in legal cases, risking wrongful convictions.

What to do if you are victim of a deepfake Scam

If you suspect a scam:

- Immediately inform your bank
- Block your cards or accounts if required
- Report to the cybercrime reporting portal
- Change your passwords and secure accounts
- Save evidence like call logs or messages



Reporting Agencies:

- **The Vanuatu Police Force (VPF) Cybercrime Unit**

Key Contact Information:
Location: Equity House, Port Vila
Main Phone (Police HQ): +678 22222
Emergency Lines: 111 (Digicel) / 1111 (Vodafone)
Email: vpfmedia@vanuatu.gov.vu
Facebook page: Vanuatu Police Force
Website: vpf.gov.vu

- **CERT Vanuatu (CertVu) report incident website or**

any information on cyber security awareness:

Phone: +678 33380

Email: info@cert.gov.vu

Link: cert.gov.vu

Messenger: CERT Vanuatu – CERT VU