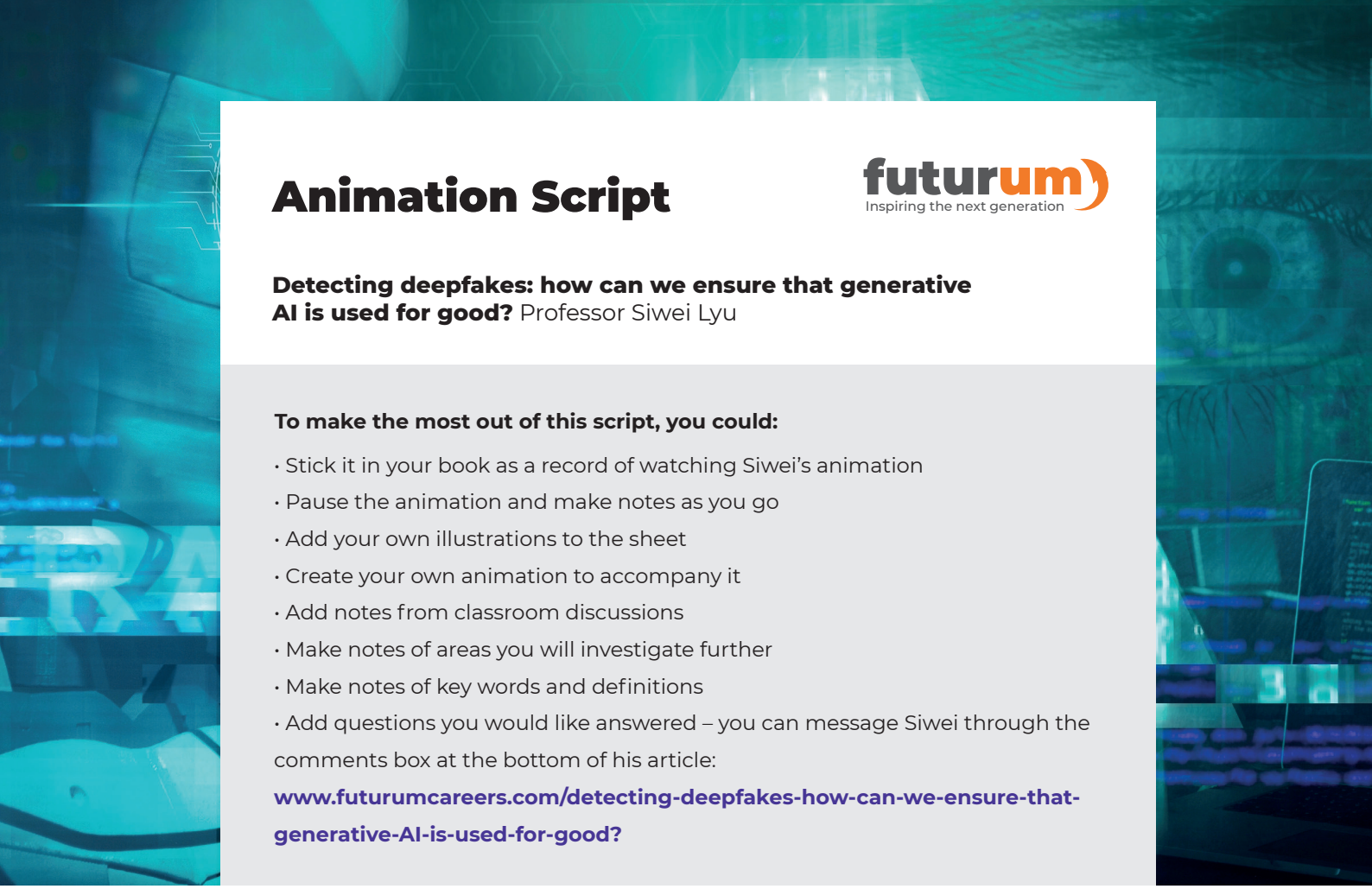

Animation Script

futurum
Inspiring the next generation

Detecting deepfakes: how can we ensure that generative AI is used for good? Professor Siwei Lyu

To make the most out of this script, you could:

- Stick it in your book as a record of watching Siwei's animation
- Pause the animation and make notes as you go
- Add your own illustrations to the sheet
- Create your own animation to accompany it
- Add notes from classroom discussions
- Make notes of areas you will investigate further
- Make notes of key words and definitions
- Add questions you would like answered – you can message Siwei through the comments box at the bottom of his article: **www.futurumcareers.com/detecting-deepfakes-how-can-we-ensure-that-generative-AI-is-used-for-good?**

SCRIPT:

Generative artificial intelligence (or generative AI) is advancing quickly, and deepfakes – manipulated pieces of media using generative AI technology – are becoming more convincing and problematic.

Professor Siwei Lyu, at the University at Buffalo, The State University of New York, in the US, is determined to halt the advance of deepfake media and ensure that generative AI is used for the good of society.

AI refers to algorithms and systems that can learn, predict and, in some cases, create. Algorithms with this creative ability are known as generative AI and can be used to fabricate media such as images, videos and audio.

Generative AI can now learn by analysing photos, audio and videos that are widely available on the internet, making it cheaper and easier for users to create convincing fake media.

Deepfakes pose personal security risks. False representations of individuals can lead to reputational damage and psychological distress.

Animation Script

Deepfakes can also impact our democratic processes by spreading disinformation – a particular issue around elections, when people are deciding who to vote for based on things they see and read online.

As generative AI develops, deepfakes will become more powerful but so will methods of detection.

AI models that create deepfakes are trained on enormous amounts of data, but they have no understanding of the laws of physics or how the human body works, and they often make mistakes.

For example, generative AI models are often trained on thousands of images of human faces downloaded from the internet. Almost all the people in these images have their eyes open. As a result, the simulated people in many deepfake videos do not blink. Other errors to look out for are missing teeth when someone is talking, hands that have the wrong number of fingers, and the reflections in eyes pointing in different directions.

Siwei has developed detection tools that, like X-ray scanners, can see inside deepfakes and uncover visual errors invisible to the human eye.

He has also developed pre-emptive tools that add specially designed patterns to 'poison' the training data, disrupt the AI's training process and cause the AI models to create low-quality deepfakes.

Generative AI can also be used for good. For example, an AI algorithm is being used to help stroke patients who have lost their ability to speak by translating the patient's brain activity into simulated speech that resembles their old voice.

Developing efficient and robust forensic methods will halt the progress of deepfake technologies and ensure that generative AI is used for societal good.

What could you achieve as a media forensics researcher?