

[illegible]

```
class SampleDataGenerator:
    @staticmethod
```

```

def create_sample_newspaper_image():
    """Create a sample newspaper image with text"""
    # Create a new image with a white background
    width = 800
    height = 600
    image = Image.new('RGB', (width, height), 'white')
    draw = ImageDraw.Draw(image)

    # Add sample text
    sample_text = """
DAILY NEWS

CITY PLANS NEW PARK
Local government announces plans for
a new community park in downtown area.
The project will begin next month.

COMMUNITY EVENTS
Senior center hosts weekly reading club
every Tuesday at 2 PM. All are welcome
to join and participate.

WEATHER REPORT
Sunny skies expected this weekend with
temperatures reaching 75 degrees.
Perfect weather for outdoor activities.
"""

    # Draw text on image
    draw.text((50, 50), sample_text, fill='black')

    # Save image to bytes
    img_byte_arr = io.BytesIO()
    image.save(img_byte_arr, format='PNG')
    img_byte_arr.seek(0)

    return img_byte_arr

class NewspaperReader:
    def __init__(self):
        self.engine = None
        self.setup_tts_engine()

    def setup_tts_engine(self):
        """Initialize TTS engine with fallback options"""
        try:
            self.engine = 'espeak' # Use espeak directly in Colab
        except:
            raise RuntimeError("Could not initialize TTS engine")

    def process_image(self, image_bytes):
        """Extract text from newspaper image using OCR"""
        try:
            image = Image.open(image_bytes)
            text = pytesseract.image_to_string(image)
            return self.clean_text(text)
        except Exception as e:
            return f"Error processing image: {str(e)}"

    def clean_text(self, text):
        """Clean and format extracted text"""
        text = re.sub(r'\s+', ' ', text)
        text = re.sub(r'^\w\s.,!?-]', ' ', text)
        text = re.sub(r'([A-Z]{2,})\s', r'\1. ', text)
        return text.strip()

    def text_to_speech(self, text, output_path=None):
        """Convert text to speech using espeak"""
        try:
            if output_path:
                subprocess.run(['espeak', '-w', output_path, text])
                return "Success"
            else:
                subprocess.run(['espeak', text])
                return "Success"
        except Exception as e:
            return f"Error converting text to speech: {str(e)}"

    def process_newspaper(self, image_bytes, output_path=None):
        """Main function to process newspaper and convert to speech"""
        if output_path:
            Path(output_path).parent.mkdir(parents=True, exist_ok=True)

```

```

    text = self.process_image(image_bytes)
    if text.startswith("Error"):
        return text, None

    result = self.text_to_speech(text, output_path)
    return result, text

def main():
    try:
        # Generate sample data
        print("Generating sample newspaper image...")
        sample_generator = SampleDataGenerator()
        image_bytes = sample_generator.create_sample_newspaper_image()

        # Create output directory
        output_dir = Path("output")
        output_dir.mkdir(exist_ok=True)

        # Process sample image
        print("\nProcessing sample newspaper...")
        reader = NewspaperReader()
        result, extracted_text = reader.process_newspaper(
            image_bytes,
            "output/sample_article.wav"
        )

        print("\nExtracted Text:")
        print("-" * 50)
        print(extracted_text)
        print("-" * 50)

        print("\nTTS Result:", result)

        # Display audio if successful
        if result == "Success":
            print("\nPlaying audio... (If you're in Colab, you can play the audio file below)")
            display(Audio("output/sample_article.wav"))

    except Exception as e:
        print(f"Error: {str(e)}")
        print("Please ensure all required dependencies are installed:")
        print("1. For OCR: !apt-get install tesseract-ocr")
        print("2. For TTS: !apt-get install espeak")
        print("3. Python packages: !pip install pytesseract pillow pytsx3")

if __name__ == "__main__": # Corrected the variable name to '__name__'
    main()

```

Generating sample newspaper image...

Processing sample newspaper...

Extracted Text:

-----

DAILY. NEWS. CITY. PLANS. NEW. PARK. Local gavernmentannounces plans for a new community park in downtown area, The project begin ne

-----

TTS Result: Success

Playing audio... (If you're in Colab, you can play the audio file below)

0:00 / 0:26

Start coding or [generate](#) with AI.

