In [87]:
```
import cv2
import re
from matplotlib import pyplot as plt
import numpy as np
import pytesseract
import pymongo
from PIL import Image
im_file = "C:/Users/neeta/temp/nvc.jpg"
img = cv2.imread(im_file)
```

In [88]:
```
def display(im_path):
    dpi = 80
    im_data = plt.imread(im_path)
    height, width = im_data.shape[:2]
    figsize = width / float(dpi), height / float(dpi)
    fig = plt.figure(figsize=figsize)
    ax = fig.add_axes([0, 0, 1, 1])
    ax.axis('off')
    ax.imshow(im_data, cmap='gray')
    plt.show()
display(im_file)
```

भारत निर्वाचन आयोग
ELECTION COMMISSION OF INDIA
मतदाता फोटो पहचान पत्र - Elector Photo Identity Card

URO3772860

नाम: नील द्विवेदी
Name: Neil Dwivedi
पिता का नाम: अनिल कुमार द्विवेदी
Father's Name: Anil Kumar Dwivedi
लिंग / Gender: पुरुष / Male
जन्म तिथि / आयु:
Date of Birth / Age: 16-11-2004

In [89]:
```
def grayscale(image):
    return cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
```

In [90]:
```
gray_image = grayscale(img)
cv2.imwrite("C:/Users/neeta/temp/graya2.jpg", gray_image)
```

Out[90]:
```
True
```

In [91]:
```
display("C:/Users/neeta/temp/graya2.jpg")
```

भारत निर्वाचन आयोग
ELECTION COMMISSION OF INDIA
मतदाता फोटो पहचान पत्र - Elector Photo Identity Card

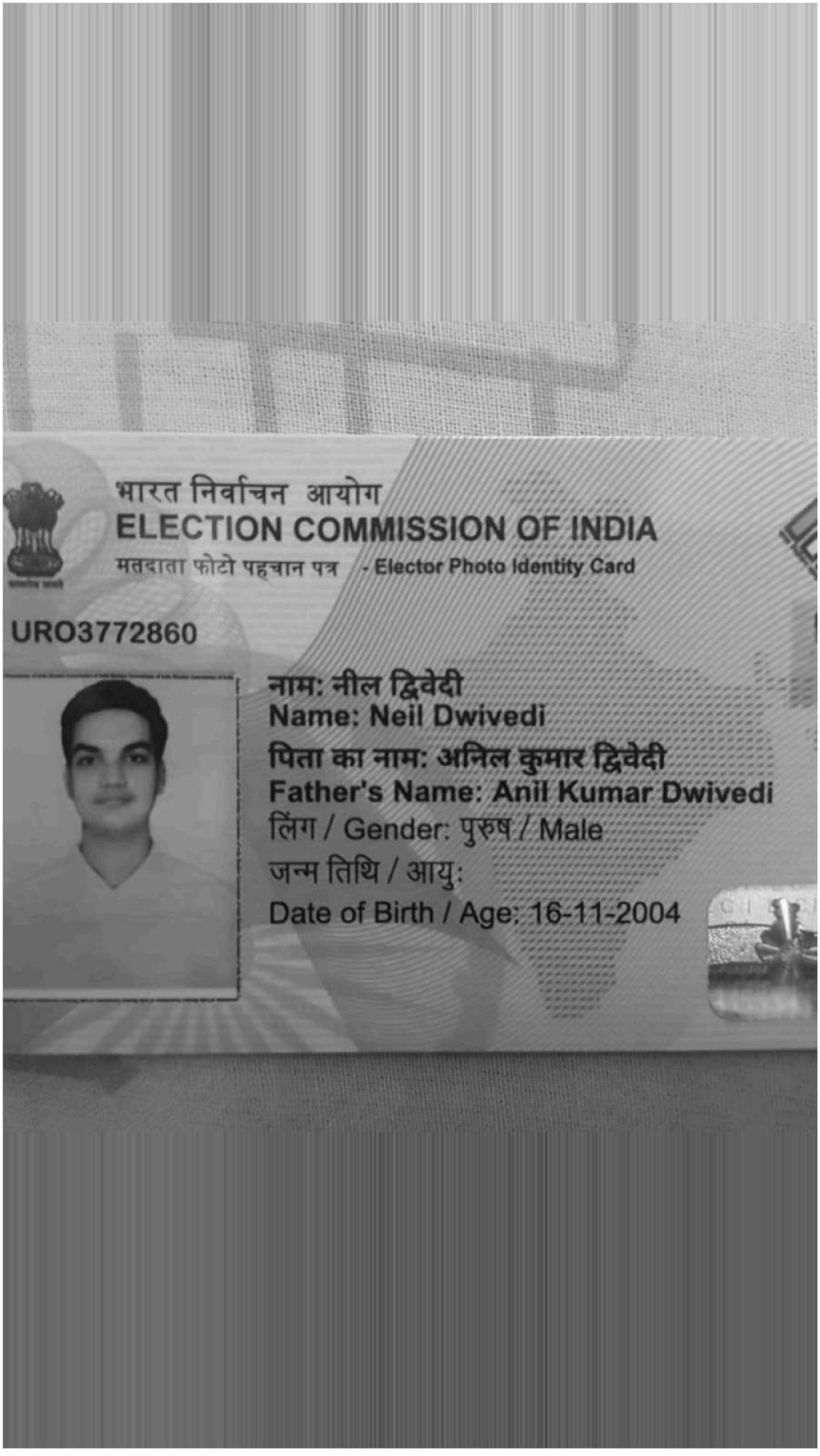
URO3772860

नाम: नील द्विवेदी
Name: Neil Dwivedi
पिता का नाम: अनिल कुमार द्विवेदी
Father's Name: Anil Kumar Dwivedi
लिंग / Gender: पुरुष / Male
जन्म तिथि / आयु:
Date of Birth / Age: 16-11-2004

In [1]:
```
def getSkewAngle(cvImage) -> float:
    newImage = cvImage.copy()
    blur = cv2.GaussianBlur(newImage, (9, 9), 0)
    thresh = cv2.threshold(blur, 0, 255, cv2.THRESH_BINARY_INV + cv2.THRESH_OTSU)[1

    kernel = cv2.getStructuringElement(cv2.MORPH_RECT, (30, 5))
    dilate = cv2.dilate(thresh, kernel, iterations=5)

    contours, hierarchy = cv2.findContours(dilate, cv2.RETR_LIST, cv2.CHAIN_APPROX_
    contours = sorted(contours, key = cv2.contourArea, reverse = True)

    largestContour = contours[0]
    minAreaRect = cv2.minAreaRect(largestContour)

    angle = minAreaRect[-1]
    if angle < -45:
        angle = 90 + angle
    return -5.0 * angle
```

In [2]:
```
def rotateImage(cvImage, angle: float):
    newImage = cvImage.copy()
    (h, w) = newImage.shape[:2]
    center = (w // 2, h // 2)
    M = cv2.getRotationMatrix2D(center, angle, 1.0)
    newImage = cv2.warpAffine(newImage, M, (w, h), flags=cv2.INTER_CUBIC, borderMod
    return newImage

def deskew(cvImage):
    angle = getSkewAngle(cvImage)
    return rotateImage(cvImage, -1.0 * angle)
```

In [94]:
```
fixim = deskew(gray_image)
cv2.imwrite("C:/Users/neeta/temp/rotfix.jpg", fixim)
display("C:/Users/neeta/temp/rotfix.jpg")
```


भारत निर्वाचन आयोग
ELECTION COMMISSION OF INDIA
मतदाता फोटो पहचान पत्र - Elector Photo Identity Card
URO3772860
नाम: नील द्विवेदी
Name: Neil Dwivedi
पिता का नाम: अनिल कुमार द्विवेदी
Father's Name: Anil Kumar Dwivedi
लिंग / Gender: पुरुष / Male
जन्म तिथि / आयु:
Date of Birth / Age: 16-11-2004

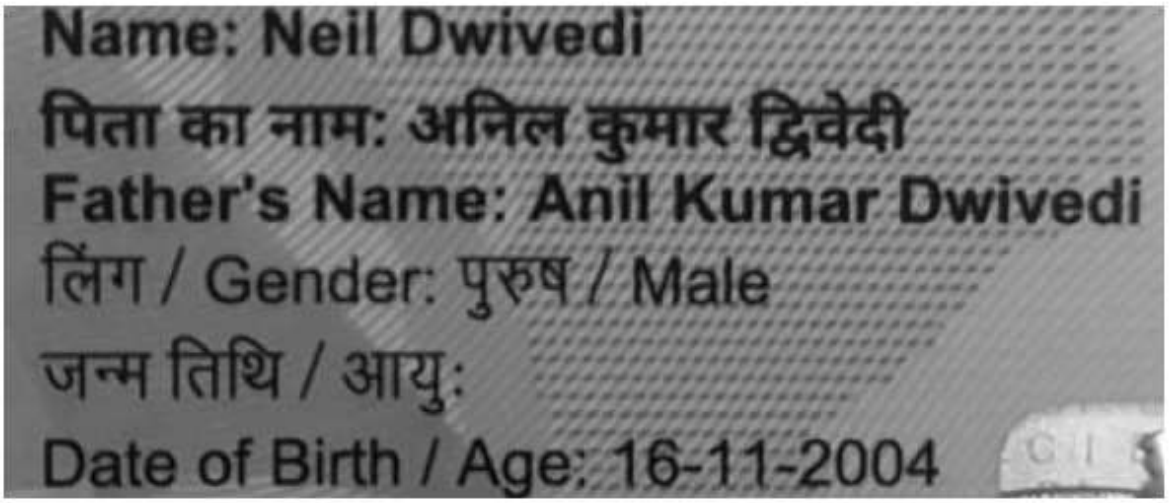
In [95]:
```
print(img.shape)
```

```
(1280, 720, 3)
```

In [96]:
```
cropped_image1 = fixim[620:820, 220:690]
cv2.imwrite("C:/Users/neeta/temp/CropAd.jpg", cropped_image1)
```

Out[96]:
```
True
```

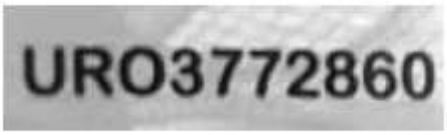
In [97]:
```
display("C:/Users/neeta/temp/CropAd.jpg")
```



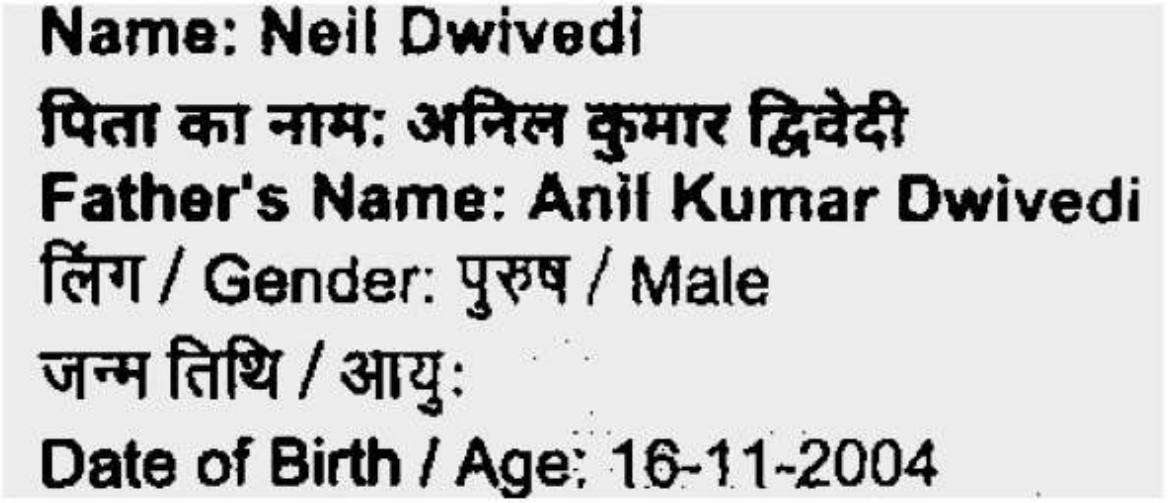
In [98]:
```
cropped_image2 = fixim[530:580, :175]
cv2.imwrite("C:/Users/neeta/temp/CropAdd.jpg", cropped_image2)
```

Out[98]:
```
True
```

In [99]:
```
display("C:/Users/neeta/temp/CropAdd.jpg")
```



In [105…
```
thresh, im_bww1 = cv2.threshold(cropped_image1, 90, 100, cv2.THRESH_BINARY)
cv2.imwrite("C:/Users/neeta/temp/bww1.jpg", im_bww1)
display("C:/Users/neeta/temp/bww1.jpg")
```



In [107…
```
thresh, im_bww2 = cv2.threshold(cropped_image2, 98, 100, cv2.THRESH_BINARY)
cv2.imwrite("C:/Users/neeta/temp/bww2.jpg", im_bww2)
display("C:/Users/neeta/temp/bww2.jpg")
```

In [108…
```
im_file = "C:/Users/neeta/temp/bww1.jpg"
```

In [109…
```
img = Image.open(im_file)
ocr_resultf = pytesseract.image_to_string(img)
```

In [110…
```
print(ocr_resultf)
```

```
Name: Neil Dwivedi

fen ar art: orfrat epee feat
Father's Name: Anil Kumar Dwivedi
fer / Gender: $89 / Male

aH fete / any:

Date of Birth / Age: 16-11-2004
```

In [111…
```
im_fileb = "C:/Users/neeta/temp/bww2.jpg"
```

In [113…
```
imge = Image.open(im_fileb)
ocr_resultg = pytesseract.image_to_string(imge)
```

In [114…
```
print(ocr_resultg)
```

```
URO3772860
```

In [115…
```
pattern = re.compile('\d.*\d')
matches = pattern.finditer(ocr_resultf)
```

In [116…
```
for match in matches:
    print(match)
```

```
<re.Match object; span=(98, 100), match='89'>
<re.Match object; span=(146, 156), match='16-11-2004'>
```

In [154…
```
pattern = re.compile('(Male|..Male)\s')
matches = pattern.finditer(ocr_resultf)
```

In [155…
```
for match in matches:
    print(match)
```

```
<re.Match object; span=(101, 108), match='/ Male\n'>
```

In [166…
```
pattern = re.compile('\w.+\s')
matches = pattern.finditer(ocr_resultf)
```

In [167…
```
for match in matches:
    print(match)
```

```
<re.Match object; span=(0, 19), match='Name: Neil Dwivedi\n'>
<re.Match object; span=(20, 49), match='fen ar art: orfrat epee feat\n'>
<re.Match object; span=(49, 83), match="Father's Name: Anil Kumar Dwivedi\n">
<re.Match object; span=(83, 108), match='fer / Gender: $89 / Male\n'>
<re.Match object; span=(109, 124), match='aH fete / any:\n'>
<re.Match object; span=(125, 157), match='Date of Birth / Age: 16-11-2004\n'>
```

In [191…
```
f1 = ocr_resultf[6:18]
f2 = ocr_resultf[64:82]
f3 = ocr_resultf[103:107]
f4 = ocr_resultf[146:156]
f5 = ocr_resultg[:10]
```

In [192…
```
print(f1)
print(f2)
print(f3)
print(f4)
print(f5)
```

```
Neil Dwivedi
Anil Kumar Dwivedi
Male
16-11-2004
URO3772860
```

In [193…
```
if __name__ == "__main__":
    client = pymongo.MongoClient("mongodb://localhost:27017")
    print(client)
    print(client.list_database_names())
```

```
MongoClient(host=['localhost:27017'], document_class=dict, tz_aware=False, connect
=True)
['OCR', 'admin', 'config', 'local', 'sample']
```

In [194…
```
    db = client['OCR']
    collection = db['VoterID']
```

In [196…
```
    dic1 = {'Name': f1, "Father's Name" : f2, 'Sex' : f3, 'D.O.B.' : f4, 'VoterID N
    collection.insert_one(dic1)
```

Out[196]:
```
InsertOneResult(ObjectId('663f420f8edc6e1db2e98c51'), acknowledged=True)
```

In [197…
```
    print(dic1)
```

```
{'Name': 'Neil Dwivedi', "Father's Name": 'Anil Kumar Dwivedi', 'Sex': 'Male', 'D.
O.B.': '16-11-2004', 'VoterID Number': 'URO3772860', '_id': ObjectId('663f420f8edc
6e1db2e98c51')}
```