

Generating Synthetic Face Data using Generative AI



Butterfly Data helped a UK police authority develop a synthetic face data generation tool using generative AI, to ethically test facial recognition (FR) technology.

Challenge

Facial recognition (FR) technology is increasingly used in policing to enhance security and operational efficiency. However, testing these systems with real-world data raises significant ethical concerns, including privacy violations and potential biases.

Our client, a UK police authority, sought a solution that would allow for comprehensive testing of FR systems without relying on actual images of individuals. The objective was to develop a method to generate synthetic facial data that accurately represents diverse demographics, enabling controlled and repeatable testing scenarios.

Solution

Butterfly Data developed a synthetic face generation tool leveraging generative AI techniques. The solution utilised an open-source model as the foundation, incorporating a custom Python-based prompt generation engine to guide the creation of realistic, diverse mugshots.

Key features of the tool included:

- Ability to specify attributes such as age, ethnicity, and gender to create varied datasets
- Techniques to reduce stereotypical representations related to clothing styles, hairstyles, and other features
- Generation of faces under different lighting conditions and expressions to test system robustness

This approach enabled the creation of large-scale, diverse synthetic datasets tailored to specific testing requirements, facilitating ethical evaluation of FR systems.

✍ More on the next page

Ready to transform your data?

Book your free discovery call and find out how our bespoke data services and solutions could help you uncover untapped potential and maximise ROI.

hello@butterflydata.co.uk
+44 29 212 00140
butterflydata.co.uk

Get in touch →

Impact

The synthetic face generation tool provided the police authority with the capability to:

- Conduct comprehensive evaluations of FR systems without using real-world images, addressing privacy concerns
- Identify potential biases in FR algorithms by testing across a wide range of demographic variables
- Inform adjustments to FR systems, enhancing accuracy and fairness in real-world applications

The project not only advanced the ethical testing of FR technology but also laid the groundwork for future research into responsible AI deployment in policing contexts.



Ready to transform your data?

Book your free discovery call and find out how our bespoke data services and solutions could help you uncover untapped potential and maximise ROI.

hello@butterflydata.co.uk
+44 29 212 00140
butterflydata.co.uk

Get in touch →

Our services



Data quality, governance, and privacy

Improve data quality, compliance and confidence in your data.



Data engineering and cloud adoption

Create scalable cloud platforms with robust data pipelines for multi-source integration.



Analytics and visualisation

Uncover actionable insights and make informed decisions.



Data Science and AI

Deploy custom machine learning models and ethical AI to optimise, automate & enhance.