Nếu bạn đang nói về "Data Synthesis" (Tổng hợp dữ liệu) thì đó là một phương pháp tạo ra dữ liệu mới từ dữ liệu hiện có, thường được sử dụng khi dữ liệu thực tế là hạn chế. Điều này khác với Data Augmentation, vì ở đây chúng ta không chỉ thay đổi mà còn tạo ra dữ liệu mới hoàn toàn. Dưới đây là một số cách thực hiện Data Synthesis:

**1. Generative Models:**

* Sử dụng mô hình sinh dữ liệu như Generative Adversarial Networks (GANs) hoặc Variational Autoencoders (VAEs) để tạo ra dữ liệu mới có cùng phân phối với dữ liệu huấn luyện.

**2. Bootstrap Sampling:**

* Lấy mẫu ngẫu nhiên từ dữ liệu hiện có để tạo ra các bản sao mới, có thể kết hợp với các biến đổi nhỏ.

**3. Function Fitting:**

* Sử dụng một hàm số để xấp xỉ dữ liệu hiện có và tạo ra dữ liệu mới dựa trên hàm số này.

**4. Mixture Models:**

* Tạo ra dữ liệu mới bằng cách kết hợp các phân phối khác nhau của dữ liệu hiện có.

**5. Resampling and Interpolation:**

* Sử dụng kỹ thuật lấy mẫu lại và nội suy để tạo ra dữ liệu mới từ các điểm dữ liệu hiện có.

**Ví dụ:**

Nếu bạn có một tập dữ liệu hình ảnh của người và bạn muốn tăng cường dữ liệu, bạn có thể sử dụng Generative Adversarial Networks (GANs) để tạo ra ảnh mới của người dựa trên những đặc điểm học được từ tập dữ liệu hiện có. Điều này sẽ giúp bạn có một tập dữ liệu lớn hơn và đa dạng hơn về các biến thể của đối tượng.

Tuy nhiên, cần lưu ý rằng việc tạo ra dữ liệu mới cũng đặt ra một số thách thức, như đảm bảo tính thực tế của dữ liệu và tránh tình trạng quá mức tinh chỉnh mô hình dựa trên dữ liệu tổng hợp mà không đảm bảo tính tổng quát.