

# IT 트렌드 리포트 Vol.7

*Growth Hackers*

editor 정유진, 이현아

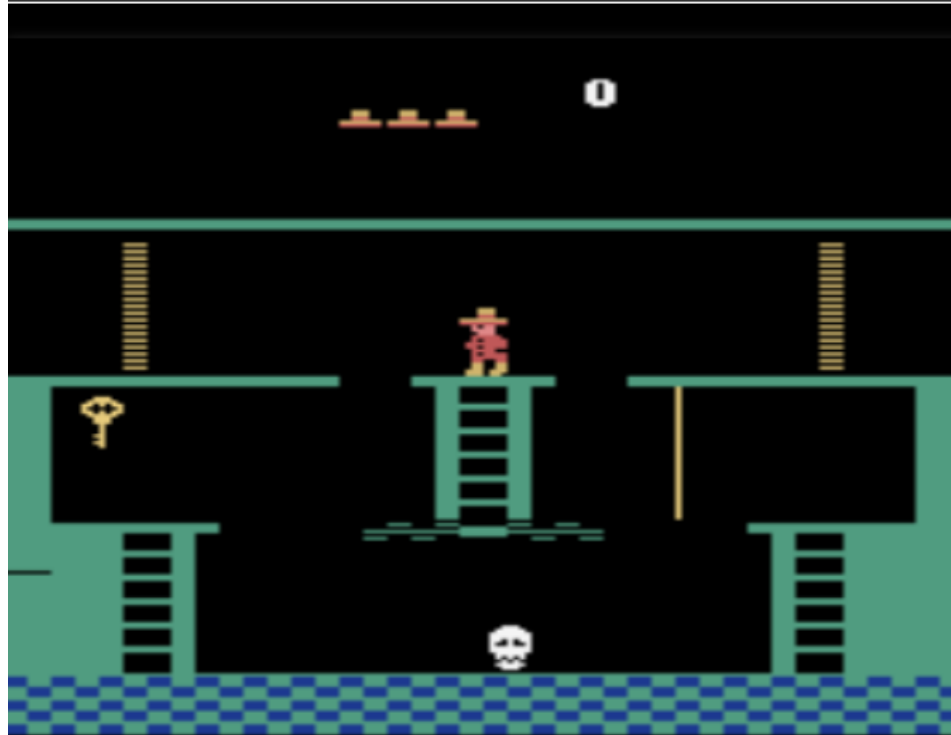


# 1

인간의 한계를 넘어 게임을 정복하다!

**Prediction-Based Rewards**  
**Reinforcement Learning**

# Montezuma's Revenge - Ata

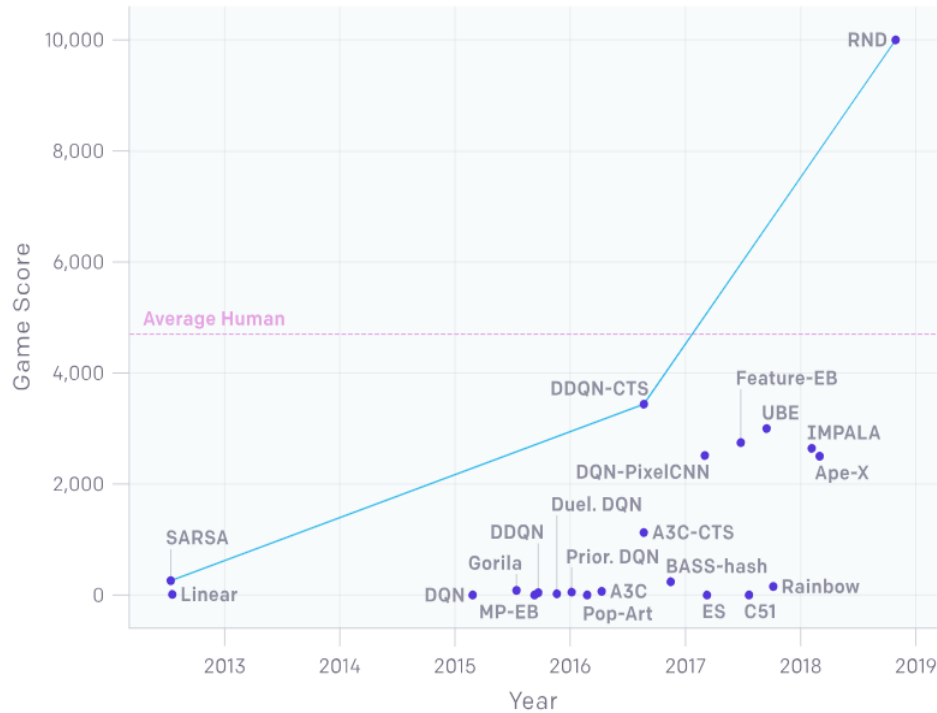


Montezuma's Revenge 게임에서 처음으로 인간의 능력을 뛰어넘는 RND 예측기반 강화학습이 나왔다.

(게임 사이트 주소

[https://www.retrogames.cz/play\\_124-Atari2600.php?language=EN/](https://www.retrogames.cz/play_124-Atari2600.php?language=EN/))

Progress in Montezuma's Revenge

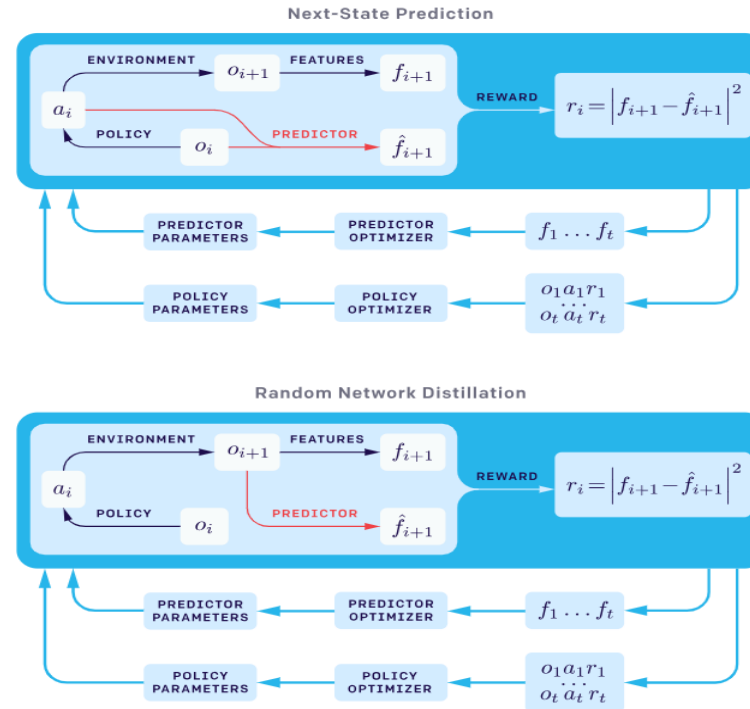


Montezuma's Revenge은 24개의 방을 찾아야하는 게임이다.

사실 이 게임은 구글 딥마인드가 개발한 딥러닝 인공지능 알고리즘인 DQN을 이용했을 때, 인간의 평균 점수(4.7K)의 0%를 받은 유일한 게임이었다.

RND로 게임에서 1레벨을 통과하고 24개의 방을 모두 찾을 수 있었다.

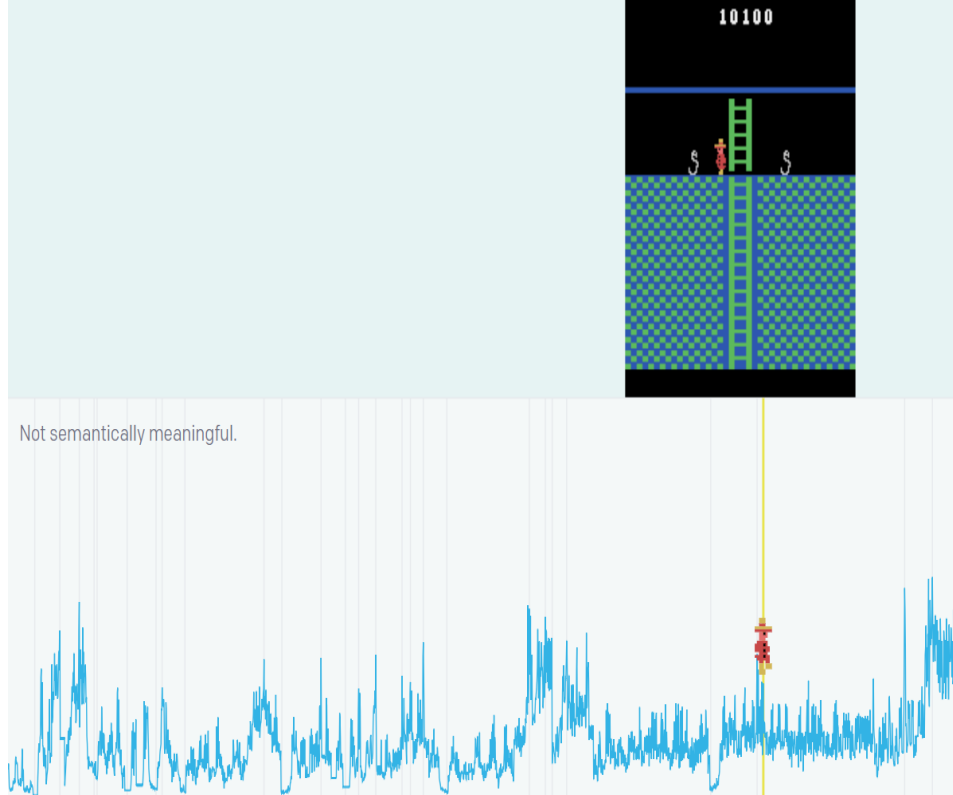
## Comparison of Next-State Prediction with RND



### RND(Random Network Distillation)의 원리

RND는 익숙하지 않은 state에서는 output을 추측하기 힘들기 때문에, Unfamiliar states에서는 더 많은 reward를 준다.

호기심에 힘입어 새로운 방을 발견하고 게임 내 점수를 높이는 방법을 찾아 훈련 후 다시 방을 다시 방문하도록 한다.



기존 연구와 다르게 관련 있느냐 없느냐에  
관계 없이 모든 환경을 모델링하지 않았다.  
단지 관찰된 특징을 고르게 했을 뿐인데  
놀랍게도 random features도 잘 작동되었다.

또한 보상을 얻는 게임이 아닌 곳에서도  
작동이 잘됐다.



Agent in a maze with a noisy TV



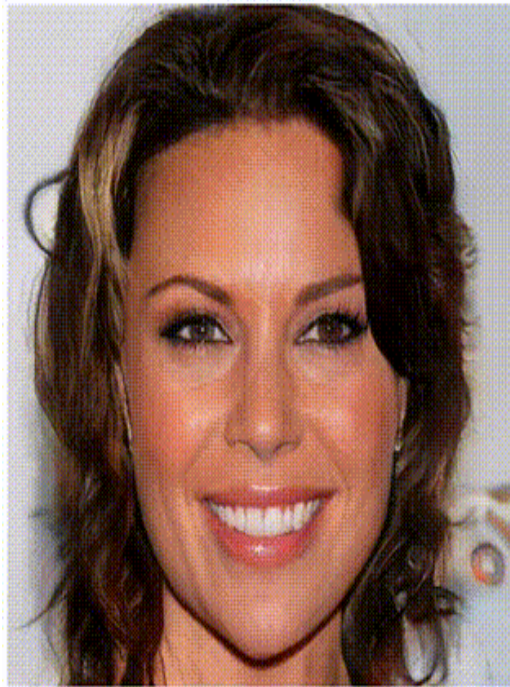
Agent in a maze without a noisy TV

**The Noisy TV Problem**: TV가 켜지면 벗어나지 않고 계속 그것을 관찰하게 되는 문제

Montezuma's Revenge에서는 탐사보너스와 게임의 점수를 합쳐 호기심의 변형을 시도.

좋은 탐사 알고리즘 선택에 The Noisy TV Problem은 매우 중요하다.

INSTRUCTION: press +/- to adjust feature, toggle feature name to lock the feature



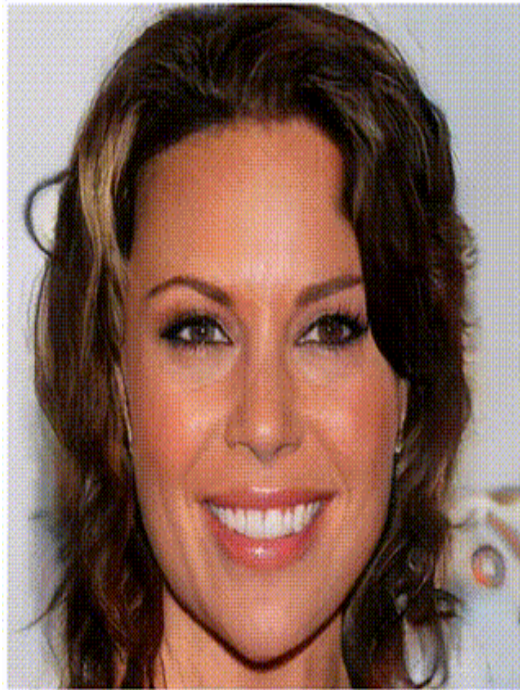
random face		
Male	Age	Skin_Tone
- +	- +	- +
Bangs	Hairline	Bald
- +	- +	- +
Big_Nose	Pointy_Nose	Makeup
- +	- +	- +
Smiling	Mouth_Open	Wavy_Hair
- +	- +	- +
Beard	Goatee	Sideburns
- +	- +	- +
Blond_Hair	Black_Hair	Gray_Hair
- +	- +	- +
Eyeglasses	Earrings	Necktie
- +	- +	- +

# 2

묘사에 맞게  
얼굴 이미지를 만들어내는 인공지능?  
**Face-generating AI**



INSTRUCTION: press +/- to adjust feature, toggle feature name to lock the feature

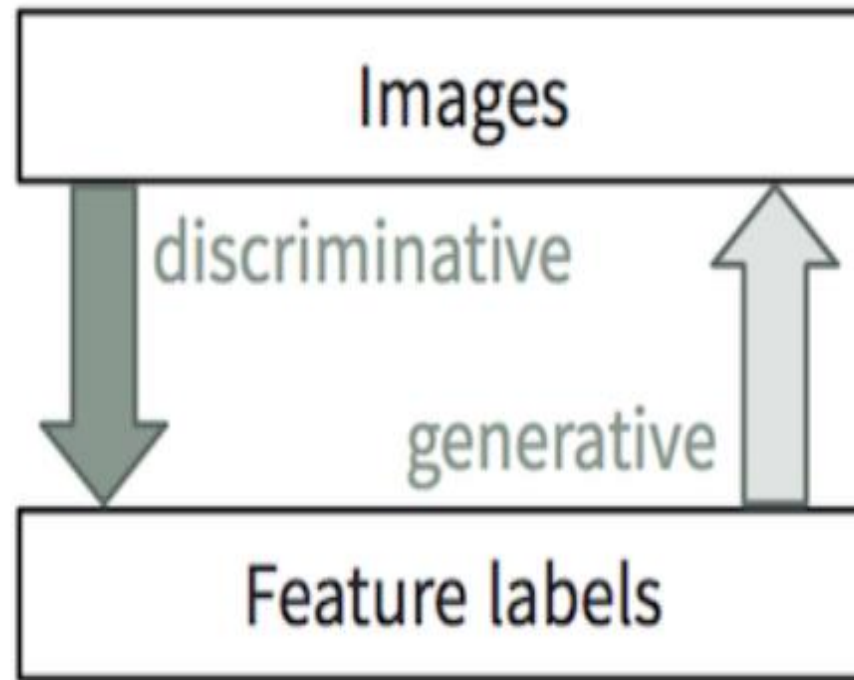


random face

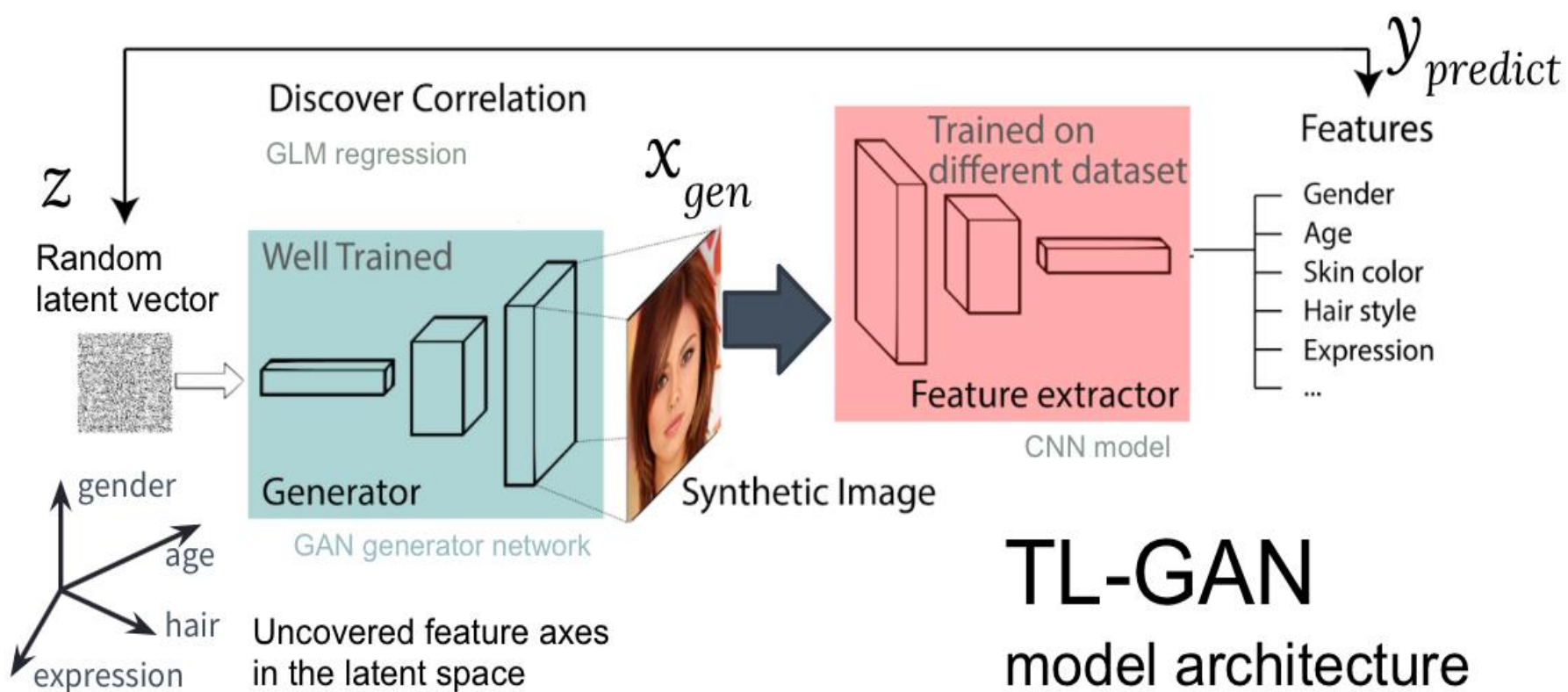
Male	Age	Skin_Tone
- +	- +	- +
Bangs	Hairline	Bald
- +	- +	- +
Big_Nose	Pointy_Nose	Makeup
- +	- +	- +
Smiling	Mouth_Open	Wavy_Hair
- +	- +	- +
Beard	Goatee	Sideburns
- +	- +	- +
Blond_Hair	Black_Hair	Gray_Hair
- +	- +	- +
Eyeglasses	Earrings	Necktie
- +	- +	- +

얼굴에 대한 묘사(descriptions) 정보를 주면  
이에 맞게 실제적인 얼굴 이미지를 만들어내는  
Face-generating AI가 발전하고 있다.

Video: <https://www.youtube.com/watch?v=O1by05eX424>

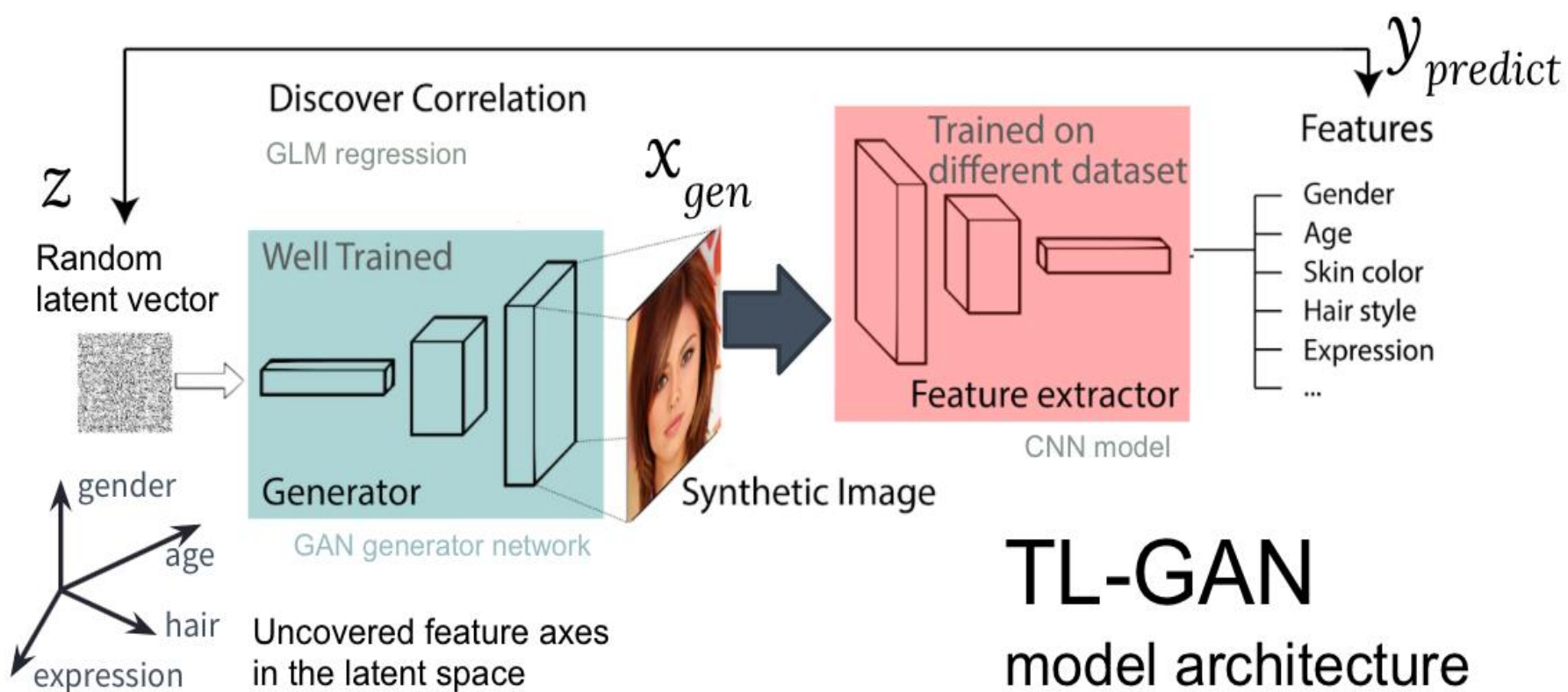


이와 같은 task는 머신 러닝에서  
**Generative task**라고 불리는데,  
이는 classification/regression과 같은  
**Discriminative task**보다 훨씬 어려운 과제이다.



최근에 고안된 TL-GAN model의 원리

- 1) Random latent vector를 input으로 하여 generator model을 통해 이미지 생성
- 2) 생성된 이미지를 input으로 하여 feature extractor model을 통해 feature 뽑아냄



최근에 고안된 TL-GAN model의 원리

- 3) 맨 처음 input인 random latent vector와 마지막 output인 features 간의 관계를 GLM regression으로 구하기
- 4) Latent vector를 조작하면서 생성된 이미지가 어떻게 변하는지 확인



## Generative model의 쓰임새는 다양하다

- 1) 콘텐츠 생성
  - ➔ 키워드에 맞는 제품 디자인 생성 가능
- 2) 쉽고 빠른 이미지 편집
- 3) Dataset 확장시키기
  - ➔ (ex) 자율주행차 회사의 training dataset





AN EXERCISE IN LANGUAGE COMPRESSION

# Are Pop Lyrics Getting More Repetitive?

By [Colin Morris](#)

# 3

재밌는 데이터 저널리즘 사례  
"Are pop lyrics getting more repetitive?"



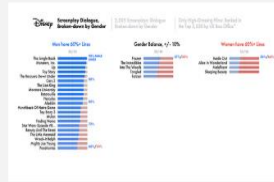
## 데이터 저널리즘이란?

데이터에 대한 분석, 시각화, 스토리화를 통해 독자들과 소통하는 것

특징1. 데이터가 보조 수단이 아니라, **그 자체로 핵심** (데이터로 인해 스토리가 나오는 것)

특징2. 데이터 수집/분석의 핵심 주체가 **정부가 아닌, 언론사와 기자**

# Greatest Hits



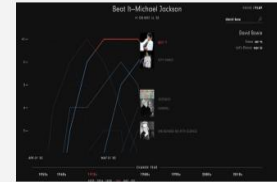
## The Largest Analysis of Film Dialogue by Gender, Ever

Matt Daniels, Hanah Anderson 2.1M views



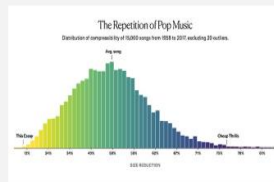
## The Largest Vocabulary in Hip Hop

Matt Daniels 1.8M views



## How Music Taste Evolved

Matt Daniels 1.4M views



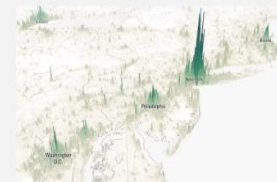
## Are Pop Lyrics Getting More Repetitive?

Colin Morris 1.1M views



## The Cultural Borders of Songs

Matt Daniels 650k views



## Human Terrain

Matt Daniels 490k views

“The Pudding”은 문화 관련 데이터 시각화를 통한 visual essay를 게시하는 사이트.

그 중 하나의 주제

“Are pop lyrics getting more repetitive?”

(사이트 주소: <https://pudding.cool/2017/05/song-repetition/>)





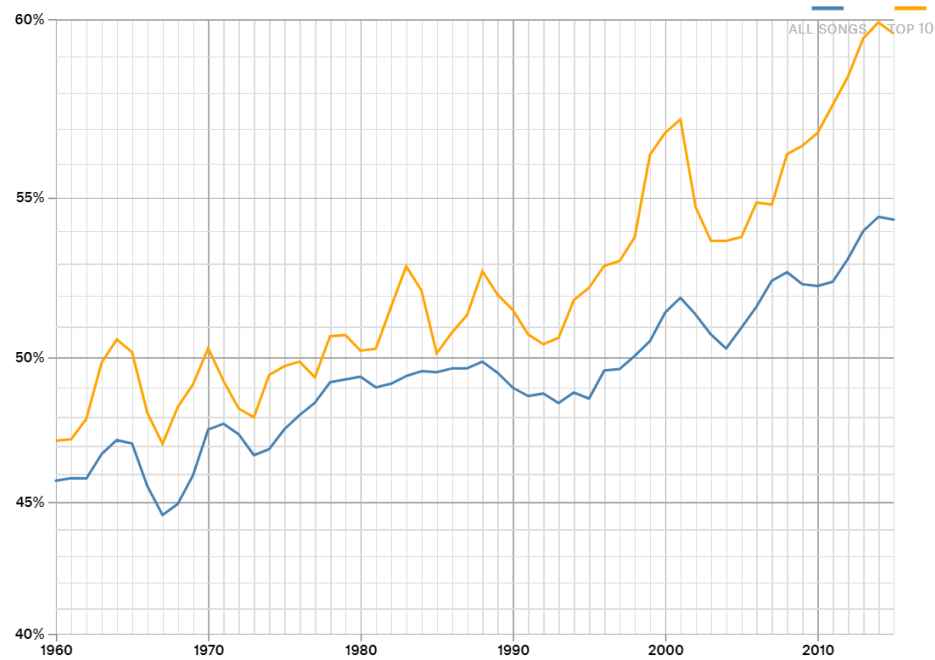
Repetitiveness  $\approx$  Compressibility

가사가 '반복적'이라는 것은,

앞의 가사와 똑같은 문자열이 나오면  
뒤에 나온 가사를 지움으로써

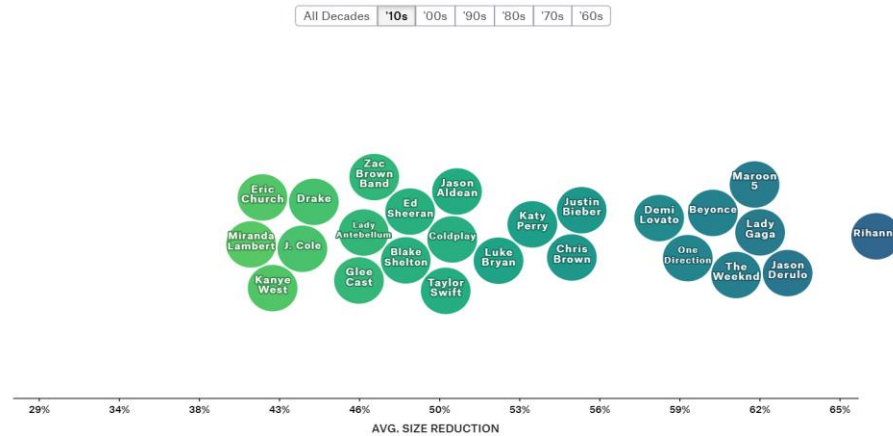
전체 가사를 얼마나 압축시킬 수 있는지  
(compressibility)와 관련 있다

## Repetition of Popular Music, by Year



그 해 발매된 노래들의 compressibility의 평균을  
기준으로 그래프를 그린 결과,  
1960년대부터 2010년대까지 상승 트렌드가  
있는 것을 알 수 있다.

## Repetitiveness Per Artist



## The Most Repetitive Songs

Of 15,000 songs from the Billboard Hot 100

All Decades '10s '00s '90s '80s '70s '60s

Track	Size Reduction
1. <b>Chain Of Fools (Part 1)</b> - Jimmy Smith, 1968	92%
2. <b>Jingo</b> - Santana, 1969	85%
3. <b>Any Way You Want It</b> - Journey, 1980	79%
4. <b>Ride!</b> - Dee Dee Sharp, 1962	81%
5. <b>You Really Got Me</b> - The Kinks, 1964	81%
6. <b>The Popcorn</b> - James Brown, 1969	81%
7. <b>My Generation</b> - The Who, 1966	80%
8. <b>Barbara Ann</b> - The Beach Boys, 1966	79%
9. <b>At My Front Door</b> - Dee Clark, 1960	78%
10. <b>I Need Your Loving</b> - Don Gardner And Dee Dee Ford, 1962	78%

- 1) 각 Artist의 평균 repetitiveness를 연도별로 시각화한 것
- 2) 각 연도별로 가장 repetitiveness가 높은 곡 10개 선정

➔ 사용자가 직접 클릭하면서 데이터를 살펴볼 수 있는 “Interactive data-visualization”



# Growth Hackers

출처:

Reinforcement Learning with Prediction-Based Rewards

- <https://blog.openai.com/reinforcement-learning-with-prediction-based-rewards/>

Face-generating AI

- <https://blog.insightdatascience.com/generating-custom-photo-realistic-faces-using-ai-d170b1b59255>

- <https://www.youtube.com/watch?v=O1by05eX424>

재밋는 데이터 저널리즘 사례 ("Are pop lyrics getting more repetitive?")

- <https://pudding.cool/2017/05/song-repetition/>

- <http://news.kbs.co.kr/news/view.do?ncd=3028270>