



**You're VAE-GAN too?**

**Generating and  
evaluating synthetic  
health data**

**Euan Gardner  
Senior Information  
Development Manager**

# Why is the NHS interested in synthetic data?

- No data-driven innovation without data
- Need for data balanced with need for privacy
- Accelerate research

# What is good synthetic data?

- Realistic
- Statistically similar
- Semantically sound

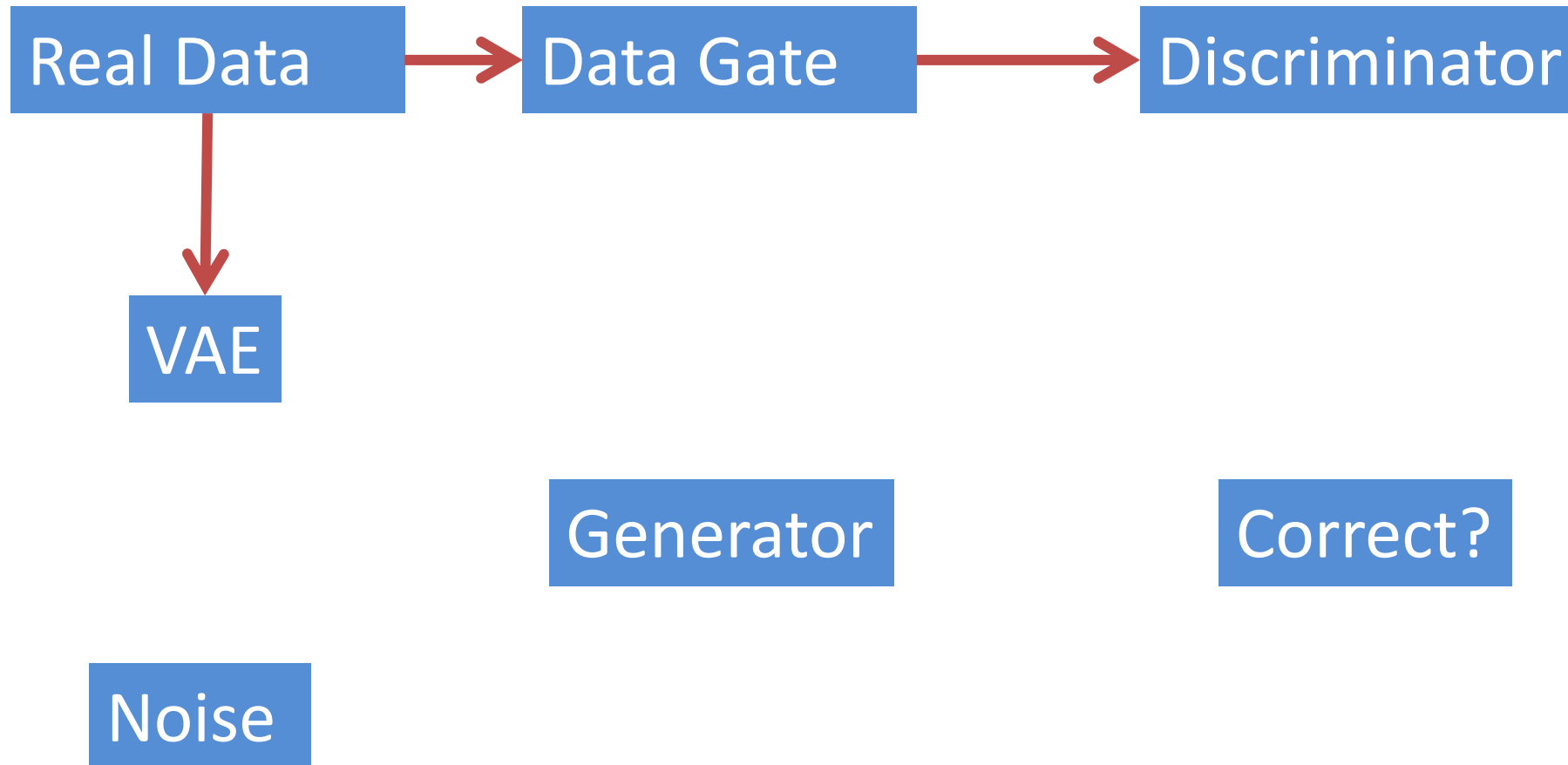


Source: <https://www.dailydot.com/unclick/bear-grylls-memes-adapt-improvise-overcome/>

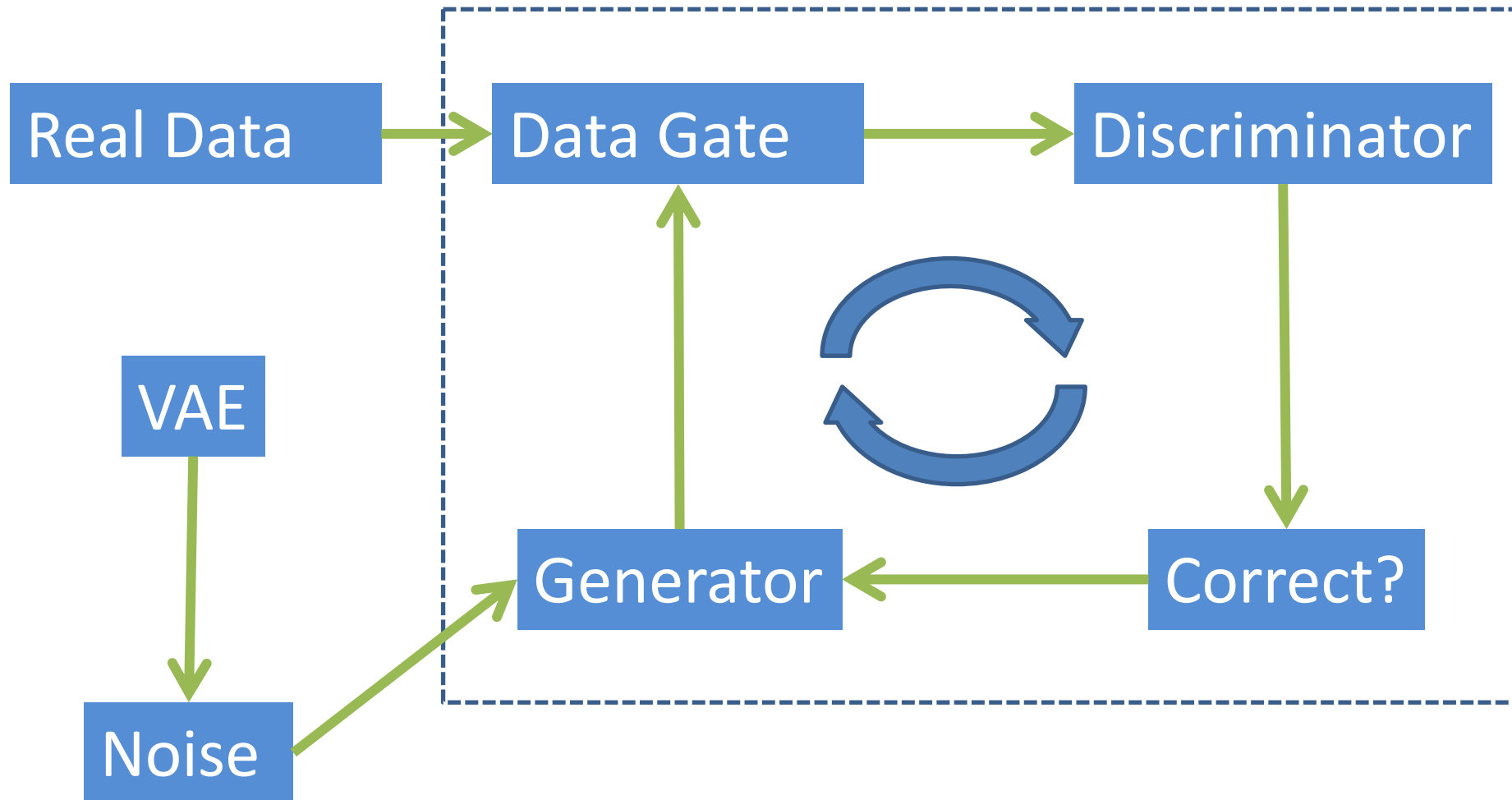
# The technology

- Artificial Neural Network approximate functions
- Variational Auto Encoder (VAE)
  - Allows differentiable variables
- Generative Adversarial Networks (GAN)
  - Powerful generation

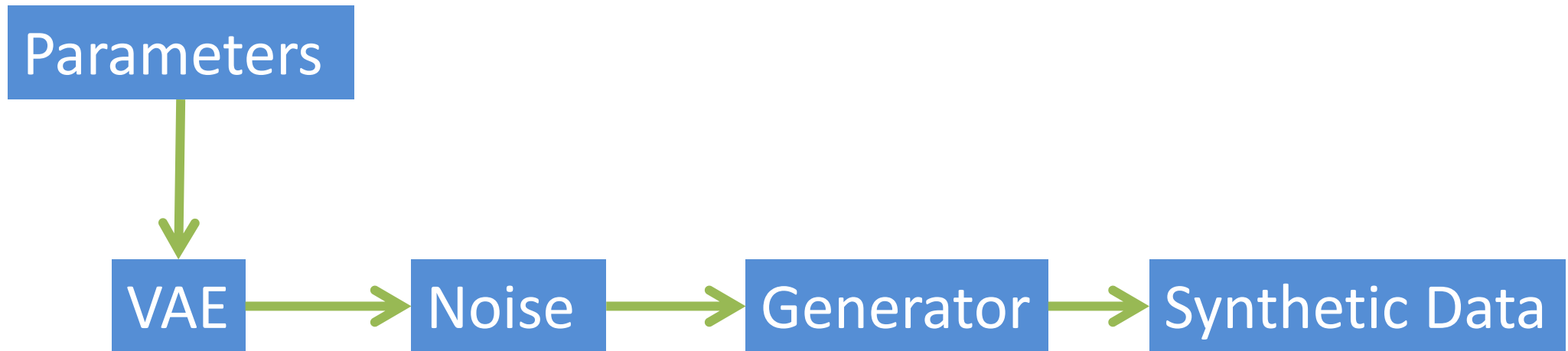
# Model: Real data training



# Model: Generator training



# Model: Synthetic data





# Where next?

- Inpatient and daycase synthesis from NHS Scotland records
- Developing synthetic data metric toolkit
- Validate outputs
- Support from The Data Lab



**Any questions?**

**Euan Gardner**  
**@SwearyStats**