There is no AI without Deep Generative Modelling

Jakub M. Tomczak

VU VRIJE UNIVERSITEIT AMSTERDAM

INFORMATION, INTELLIGENCE AND ARTIFICIAL INTELLIGENCE

What is **intelligence**?

INFORMATION, INTELLIGENCE AND ARTIFICIAL INTELLIGENCE

What is **intelligence**?

…

INFORMATION, INTELLIGENCE AND ARTIFICIAL INTELLIGENCE

What is **intelligence**?

…

What is **artificial intelligence**?

INFORMATION, INTELLIGENCE AND ARTIFICIAL INTELLIGENCE

What is **intelligence**?

…

What is **artificial intelligence**?

INFORMATION, INTELLIGENCE AND ARTIFICIAL INTELLIGENCE

What is **intelligence**?

…

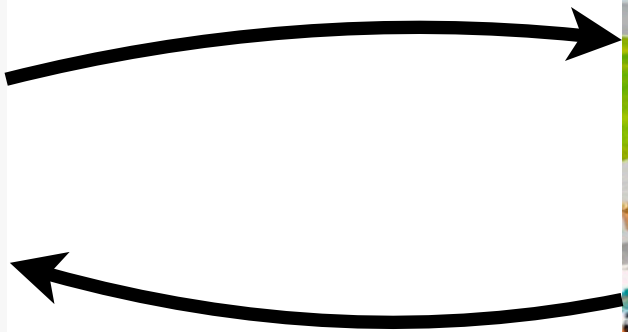
What is **artificial intelligence**?

INFORMATION, INTELLIGENCE AND ARTIFICIAL INTELLIGENCE

What is **intelligence**?

…

What is **artificial intelligence**?



INFORMATION, INTELLIGENCE AND ARTIFICIAL INTELLIGENCE

What is **intelligence**?

…

What is **artificial intelligence**?

Decision/action

Information

0001101010011…

INFORMATION, INTELLIGENCE AND ARTIFICIAL INTELLIGENCE

What is **artificial intelligence**?

- **Information** processing
- **Information** storing
- **Information** transmission

INFORMATION, INTELLIGENCE AND ARTIFICIAL INTELLIGENCE

What is **artificial intelligence**?

- **Information** processing
- **Information** storing
- **Information** transmission
- **Decision** making

INFORMATION, INTELLIGENCE AND ARTIFICIAL INTELLIGENCE

What is **artificial intelligence**?

- **Information** processing
- **Information** storing
- **Information** transmission
- **Decision** making

} **Learning**, **Knowledge representation**, **Models…**

INFORMATION, INTELLIGENCE AND ARTIFICIAL INTELLIGENCE

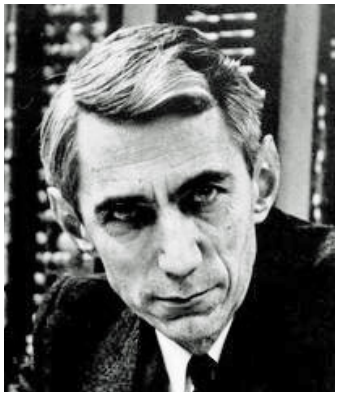
What is **artificial intelligence**?

- **Information** processing
- **Information** storing
- **Information** transmission
- **Decision** making

} **Learning**, **Knowledge representation**, **Models…**

The question is how to formalize the problem of AI?

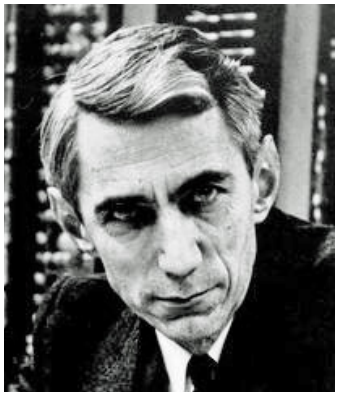
Information (a quick recap)



Claude Shannon

Information (a quick recap)

We have a random source of data x.



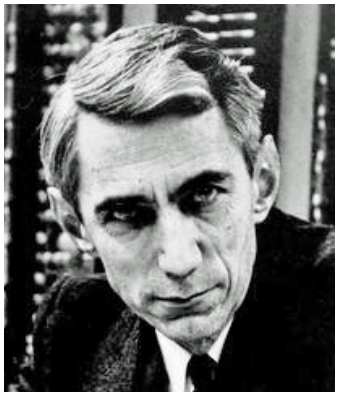
Claude Shannon

Information (a quick recap)

We have a random source of data x.

We can quantify the **uncertainty** of this source by calculating **the entropy**:

$$\mathbb{H}[x] = -\sum_{x} p(x)\log p(x)$$



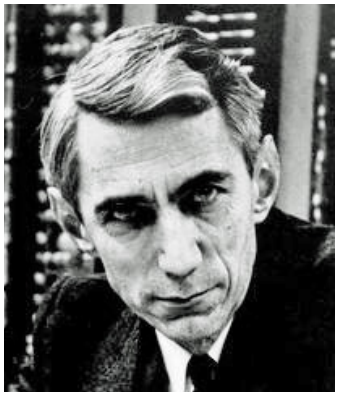
Claude Shannon

Information (a quick recap)

We have a random source of data x.

We can quantify the **uncertainty** of this source by calculating **the entropy**:

$$\mathbb{H}[x] = -\sum_{x} p(x)\log p(x)$$



Claude Shannon

Entropy is max if all x's are equiprobable.

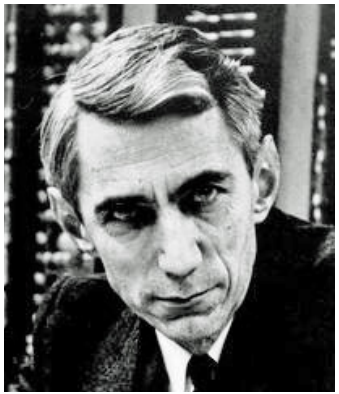
Entropy is min if the probability of one value is 1.

Information (a quick recap)

We have a random source of data x.

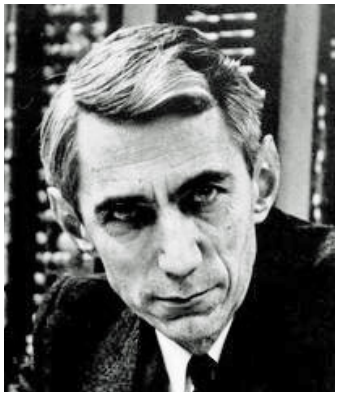
We can quantify the **uncertainty** of this source by calculating **the entropy**:

$$\mathbb{H}[x] = -\sum_{x} p(x)\log p(x)$$



Claude Shannon

Optimal message length $\approx$ the entropy.


Claude Shannon

Information (a quick recap)

We have two random sources: x and y.

We can quantify the **uncertainty** of them by calculating **the joint entropy**:

$$\mathbb{H}[x, y] = -\sum_{x,y} p(x, y) \log p(x, y)$$

or **the conditional entropy**:

$$\mathbb{H}[y \mid x] = -\sum_{x,y} p(x, y) \log p(y \mid x)$$

VU

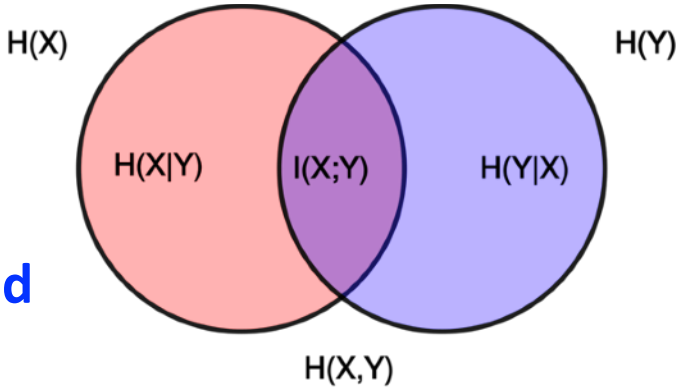
Mutual Information (a quick recap)

We have two random sources: x and y.

Mutual Information (a quick recap)

We have two random sources: x and y.

We can quantify how much **information is shared by the two sources:**

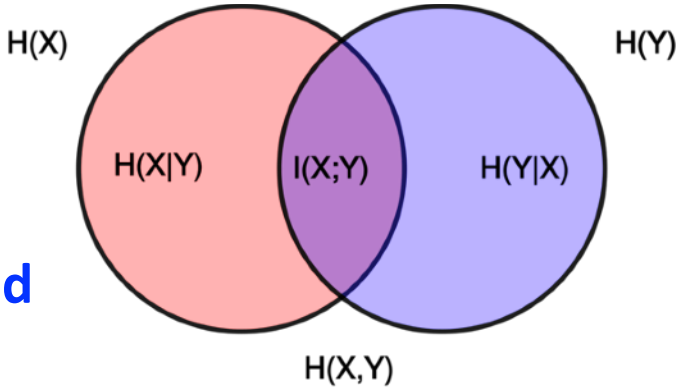


$$\mathbb{I}[x;y] = \mathbb{H}[y] - \mathbb{H}[y\,|\,x]$$

Mutual Information (a quick recap)

We have two random sources: x and y.

We can quantify how much **information is shared by the two sources:**



$$\mathbb{I}[x;y] = \mathbb{H}[y] - \mathbb{H}[y\,|\,x]$$

or **how much knowing one source reduces uncertainty about the other.**

We have two random sources: x (e.g., images) and y (e.g., decisions).

We have two random sources: x (e.g., images) and y (e.g., decisions).

We have also a **model** m (a representation of a world).

We have two random sources: x (e.g., images) and y (e.g., decisions).

We have also a **model** m (a representation of a world).

The **goal** of AI is to **maximize** the **mutual information** between (x, y) and m:

$$\mathbb{I}[(x, y); m] = \mathbb{H}[x, y] - \mathbb{H}[x, y \mid m]$$

We have two random sources: x (e.g., images) and y (e.g., decisions).

We have also a **model** m (a representation of a world).

The **goal** of AI is to **maximize** the **mutual information** between (x, y) and m:

$$\mathbb{I}[(x, y); m] = \mathbb{H}[x, y] - \mathbb{H}[x, y \mid m]$$

Entropy of the world (model has no influence on that)

That's the "real" goal!

The **goal** of AI is to **maximize** the **mutual information** between (x, y) and m

(or minimize $\mathbb{H}[x, y \mid m]$, i.e., minimize uncertainty of the world):

$$\mathbb{H}[x, y \mid m] = \sum_{x,y,m} p(x, y, m)\left[\log p(y \mid x, m) + \log p(x \mid m)\right]$$

The **goal** of AI is to **maximize** the **mutual information** between (x, y) and m

(or minimize $\mathbb{H}[x, y \mid m]$, i.e., minimize uncertainty of the world):

$$\mathbb{H}[x, y \mid m] = \sum_{x,y,m} p(x, y, m)\left[\log p(y \mid x, m) + \log p(x \mid m)\right]$$

A model for decision making

A model for understanding the world.

The **goal** of AI is to **maximize** the **mutual information** between (x, y) and m (or minimize $\mathbb{H}[x, y \mid m]$, i.e., minimize uncertainty of the world).

In order to achieve that, AI should focus on learning **two models**:

- **A model for decision making:** $p(y \mid x, m)$
- **A model for understanding the world:** $p(x \mid m)$

WHAT HAPPENS IF WE LEARN ONLY DECISION MAKING

The bulk of AI is focused on the decision making part **only**!

WHAT HAPPENS IF WE LEARN ONLY DECISION MAKING

The bulk of AI is focused on the decision making part **only**!

Example: Let's say we have a model that is well trained.

$p(y = cat|\mathbf{x}) = 0.90$
$p(y = dog|\mathbf{x}) = 0.05$
$p(y = horse|\mathbf{x}) = 0.05$

WHAT HAPPENS IF WE LEARN ONLY DECISION MAKING

The bulk of AI is focused on the decision making part **only**!

Example: Let's say we have a model that is well trained.

$p(y = cat|\mathbf{x}) = 0.90$
$p(y = dog|\mathbf{x}) = 0.05$
$p(y = horse|\mathbf{x}) = 0.05$

+

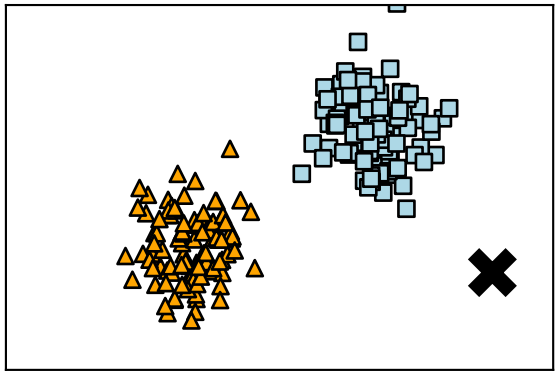
noise

=

$p(y = cat|\mathbf{x}) = 0.05$
$p(y = dog|\mathbf{x}) = 0.05$
$p(y = horse|\mathbf{x}) = 0.90$

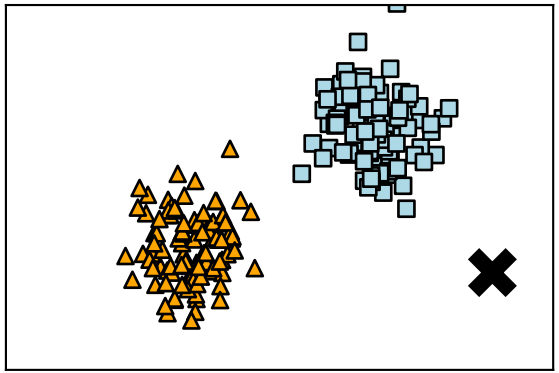
But after adding a little noise it could fail completely…

DEEP GENERATIVE MODELING: WHY DO WE NEED THEM?

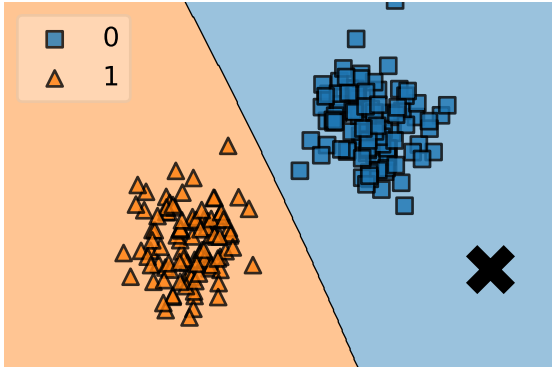


Data

DEEP GENERATIVE MODELING: WHY DO WE NEED THEM?



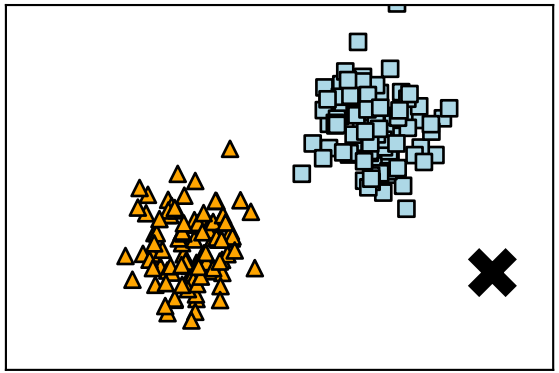
Data



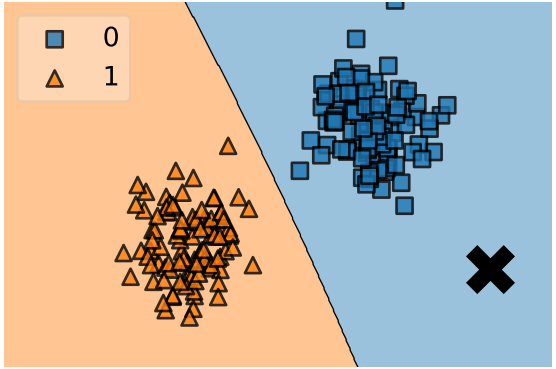
$p(y|\mathbf{x})$

$p(blue|\mathbf{x})$ is high
= certain decision!

DEEP GENERATIVE MODELING: WHY DO WE NEED THEM?

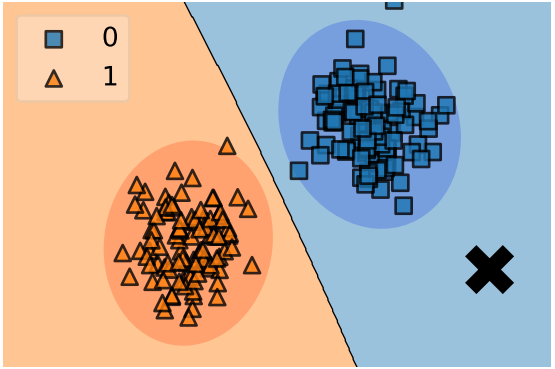


Data



$p(y|\mathbf{x})$

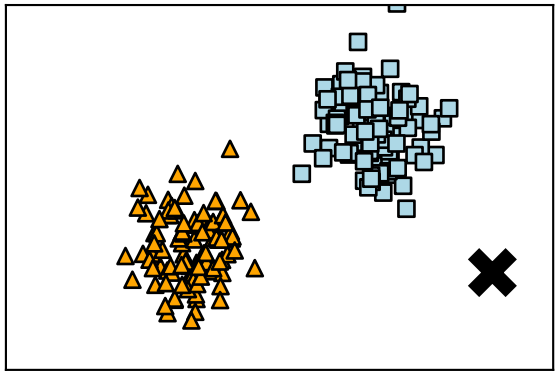
$p(blue|\mathbf{x})$ is high
= certain decision!



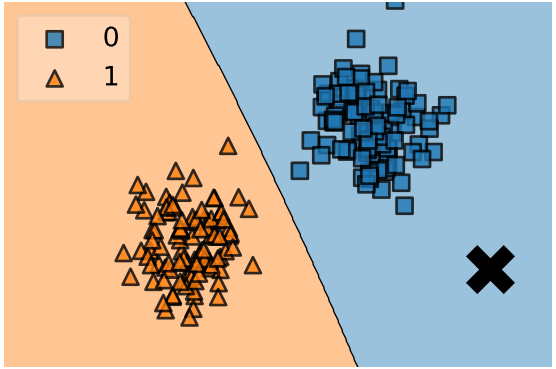
$p(\mathbf{x}, y) = p(y|\mathbf{x})\ p(\mathbf{x})$

$p(blue|\mathbf{x})$ is high
and $p(\mathbf{x})$ is low
= uncertain decision!

DEEP GENERATIVE MODELING: WHY DO WE NEED THEM?

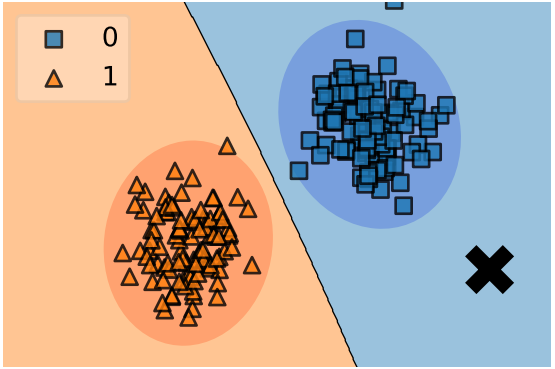


Data



$p(y|\mathbf{x})$

$p(blue|\mathbf{x})$ is high
= certain decision!



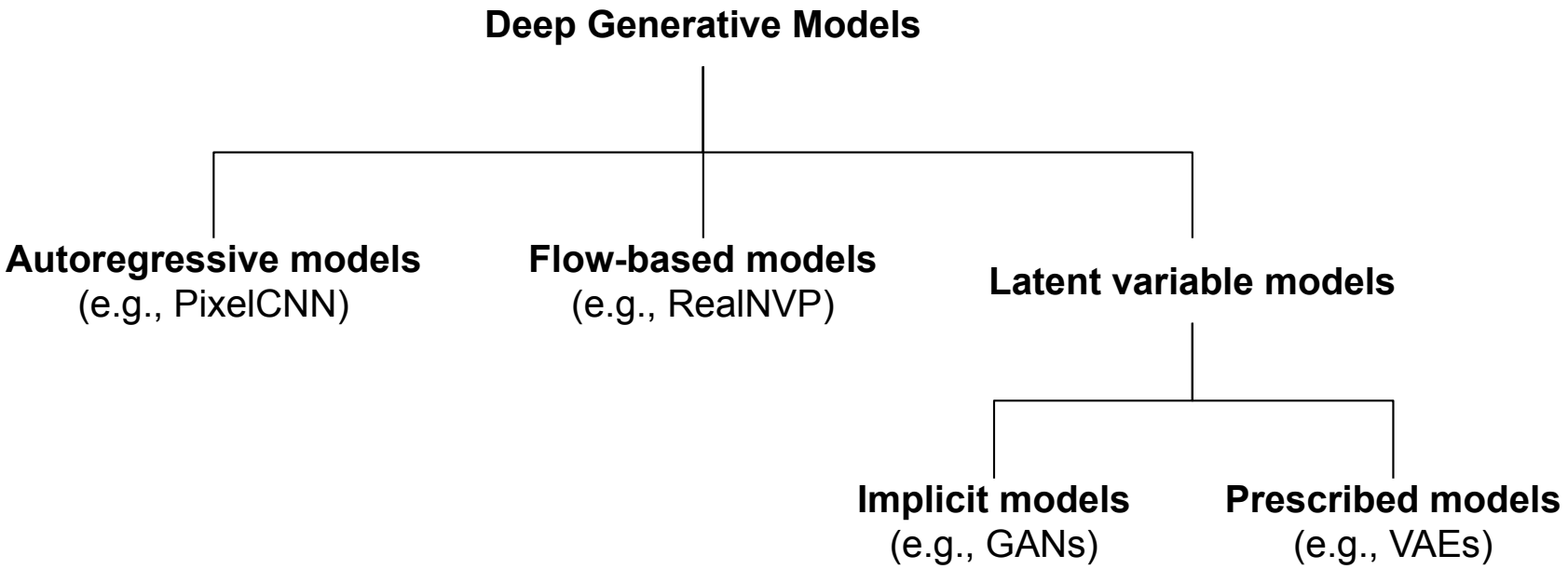
$p(\mathbf{x}, y) = p(y|\mathbf{x})\ p(\mathbf{x})$

$p(blue|\mathbf{x})$ is high
and $p(\mathbf{x})$ is low
= uncertain decision!

Thus, learning the conditional is only a part of the story!
How can we learn p(x)?

VU

DEEP GENERATIVE MODELING: HOW WE CAN FORMULATE IT?



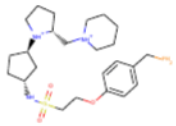
DEEP GENERATIVE MODELING: WHERE CAN WE USE IT?

“ i want to talk to you . ”
“i want to be with you . ”
“i do n’t want to be with you . ”
i do n’t want to be with you .
she did n’t want to be with him .

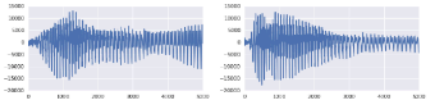
he was silent for a long moment .
he was silent for a moment .
it was quiet for a moment .
it was dark and cold .
there was a pause .
it was my turn .

Text analysis

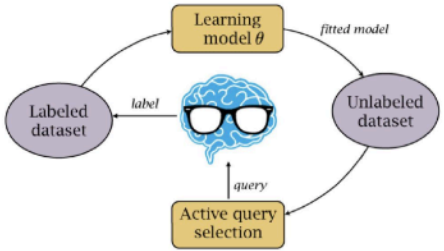
Image analysis



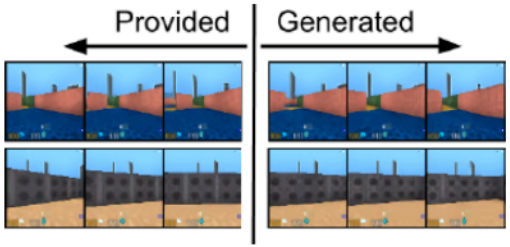
Graph analysis



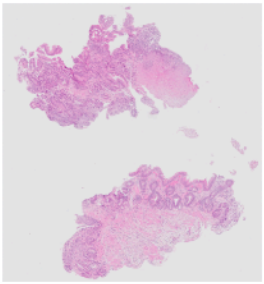
Audio analysis



Active Learning



Reinforcement Learning



Medical data

and more...

THE TAKE AWAY MESSAGE

- A decision making model is **fine** but it doesn't bring us closer to true* AI.
- **Understanding** reality and properly **quantifying uncertainty** is **crucial** in AI.

*whatever it means…

BLOG ABOUT DEEP GENERATIVE MODELING

If you are interested in going deeper into deep generative modeling, please take a look at my blog: **[Blog]**

- **Intro**: [Link]
- **ARMs**: [Link]
- **Flows**: [Link], [Link]
- **VAEs**: [Link], [Link]
- **Hybrid modeling**: [Link]

THANK YOU FOR YOUR ATTENTION

Jakub M. Tomczak
Computational Intelligence group
Vrije Universiteit Amsterdam

Webpage: https://jmtomczak.github.io/

Github: https://github.com/jmtomczak

Twitter: https://twitter.com/jmtomczak